2006

Investigation of Protein-DNA Contacts in the Saccharomyces Cerevisiae RNA Polymerase III Recruitment Factor (TFIIIB) Complex and Effects of Promoter Topology on Transcription

Nick Dimitrios Tsihlis
Louisiana State University and Agricultural and Mechanical College, ntsihl1@lsu.edu

INVESTIGATION OF PROTEIN-DNA CONTACTS IN THE *SACCHAROMYCES CEREVISIAE* RNA POLYMERASE III RECRUITMENT FACTOR (TFIIIB) COMPLEX AND EFFECTS OF PROMOTER TOPOLOGY ON TRANSCRIPTION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Nick Dimitrios Tsihlis
B.A., Rice University, 2000
December 2006

DEDICATION

To my daughter Irene, who made the last year of my graduate career the best.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Anne Grove, for her guidance, support, encouragement, and advice. Besides teaching me how to think critically about scientific questions and experimental results, she has taught me the value of choosing my words carefully, and continues to humble this native English speaker with her command of the language.

I would also like to thank my committee—Dr. John Battista, Dr. Pat DiMario, Dr. Norimoto Murai, and Dr. Grover Waldrop—for their helpful discussions and critiques. A special thanks goes out to my lab mates for making me laugh, keeping me sane, and teaching me new languages.

My deepest gratitude goes out to my family for their support and love. To my sister, Vasilia, thanks for helping me keep my perspective and telling me to hang in there. To my mother, Maria, thanks for your years of hard work and sacrifice. To my father, Dimitrios, thanks for passing on your curiosity and helping make me the person I am. To my two Springer Spaniels, Sonny and Brady, thanks for helping me find time to relax, even when it wore me out. Sonny, I will miss you.

Most of all, I would like to thank my wife, Heather, whose belief in me has never wavered, even as life brought new challenges. Her belief gives me strength, and her patience and love know no bounds. Her capacity to care for our home and our daughter, while excelling as a chemical engineer, constantly amazes me. Finally, to my daughter Irene, thanks for showing me that the spark of curiosity has been passed. You give me hope and remind me that the most important things in life happen outside the lab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The TATA-binding protein (TBP) is required for transcription by all three nuclear RNA polymerase systems. TBP binds the TATA box (consensus sequence TATAa/tAa/tN) and induces an ~80° bend in the DNA. To achieve recruitment of *Saccharomyces cerevisiae* RNA polymerase III, TBP is associated with two additional factors, Brf1 and Bdp1, to form initiation factor TFIIIB. This study focuses on interactions between the proteins and DNA in TFIIIB, and effects of promoter topology on transcription.

Previous data suggests that the structure or dynamics of the TBP-DNA complex may be altered upon entry of Brf1 and Bdp1 into the complex. Here, an altered specificity TBP mutant TBPm3 and iterative *in vitro* selection assays are used to show that Brf1 and Bdp1 impose strict sequence preference on TBPm3 for the downstream half of the TATA box. Notably, the selected sequence (TGTAAATA) perfectly matches the TATA box of the pol III-transcribed U6 small nuclear RNA (*SNR6*) gene, suggesting that the selected T•A base pair step at the downstream end of the 8 bp TBP site may provide a DNA flexure that promotes TFIIIB-DNA complex formation.

Increased TBP-induced bend angle has been correlated with increased levels of relative pol II transcription from DNA with variant TATA boxes. Here, circular permutation and electrophoretic mobility shift assays are used to show that, in a pol III system, TATA box sequence has almost no effect on protein-induced bend angle. *In vitro* transcription is used to show that, despite the lack of difference observed in TBP-induced bend angles, differences in relative transcription distinct from those observed in a pol II system occur; therefore, it may be that flexure or dynamics of the TATA box sequence governs transcription efficiency.

Notably, despite presence of a native pol III composed of 7 consecutive thymines terminator, termination is unexpectedly inefficient. Since the templates contain only five base pairs of pol III gene sequence beyond the terminator, these data suggest that sequence well beyond the terminator may affect proper termination of transcription, and that termination by pol III requires more than the well-characterized string of thymines.

CHAPTER 1

INTRODUCTION

Transcription is one of the most central processes in all organisms, converting the information encoded in DNA into RNA. In eukaryotes, three separate yet interconnected systems have evolved to produce the components necessary for life. The nucleolar RNA polymerase (pol) I is responsible for the production of ribosomal RNA (rRNA), which in turn produces proteins in the ribosome using instructions produced by RNA pol II in the form of messenger RNA (mRNA). Pol II is responsible for the production of the bulk of the cell's mRNA and small nuclear RNA (snRNA). Pol III is perhaps the most vital, as it is responsible for the production of transfer RNA (tRNA), 5S rRNA and U6 snRNA. This discussion will focus on the large amount of data that has been gathered on the yeast pol III system, making it one of the most well-characterized systems known.

Discovery of RNA Polymerase III

RNA polymerase III has been isolated from many eukaryotes, including frogs (1, 2), humans (3, 4), mice (5, 6) and yeast (7-10). Since sonication of sea urchin and rat liver nuclei at high salt concentrations released several RNA polymerases (11), similar experiments were performed on yeast nuclei and the lysates subjected to chromatography on DEAE-Sephadex (12). The elution patterns for the yeast fractions from DEAE-Sephadex are similar to those observed for sea urchin and rat, yielding 3-4 peaks with transcriptional activity (7). The proteins comprising the peaks can be further identified based on their response to α-amanitin, a potent mushroom toxin that inhibits elongation of transcripts. In higher eukaryotes, RNA pol III is completely inhibited only when the concentration is higher than 200 μg/mL (13), while pol II is inhibited at 0.5 μg/mL, and pol I is unaffected even at the highest concentration of toxin.

However, in yeast, the pol III and pol I sensitivities are reversed, with pol III completely resistant and pol I completely inhibited (14).

When pol III was first isolated from yeast, little was known about the size of the polypeptide subunits of which it is comprised. Further investigations (14) led to the elucidation of the composition of this massive complex and its similarity to other RNA polymerases. The nomenclature for RNA polymerase subunits uses letters to indicate the polymerase where the subunit is found (Pol I = A, Pol II = B, Pol III = C) and a number indicating the size of the polypeptide, so ABC27 is the 27 kDa protein found in all three polymerases, whereas B32 is the 32 kDa subunit unique to pol II. As shown in Table 1.1, the polymerases consist of large, related subunits and several smaller, shared subunits.

This makes sense evolutionarily, as the large subunits probably diverged from a common ancestor--bacterial polymerase, with which the eukaryotic polymerases are related--to specialize in the three systems, but the smaller, shared factors probably perform the same functions in all three systems. The core of bacterial RNA polymerase consists of four subunits, α, β, β', and ω, and has stoichiometry $\alpha_2\beta\beta'\omega$ (15). The core of pol III consists of: C160--a β' homolog, C128--a β homolog, and AC40 and AC19--both α homologs (16). RPB6, a protein found in all three polymerases, has been shown to be related to ω (17). A subassembly of three subunits that dissociate during chromatographic purification consists of C31, C34 and C82 (18). Figure 1.1 shows the subunit interactions determined via photochemical crosslinking in the transcribing pol III complex and their positions on the DNA.

It is energetically favorable to have the smaller subunits shared, in order to allow rapid RNA synthesis when conditions for the organism change. Pol II holoenzymes, enzyme

Table 1.1. **Subunits of *Saccharomyces cerevisiae* RNA polymerases.** Polymerase III contains 7 shared subunits, 2 related subunits, and 7 unique subunits. Subunits that make up the core of pol III are in bold. The three subunits that make up a small, dissociable complex are in italics. [a] This subunit is related to AC40.

Pol I	Pol II	Pol III
Shared subunits		
ABC10α	ABC10α	ABC10α
ABC10β	ABC10β	ABC10β
ABC14.5	ABC14.5	ABC14.5
AC19		**AC19**
ABC23	ABC23	ABC23
ABC27	ABC27	ABC27
AC40		**AC40**
Related subunits		
A12.2	B12.6	
	B44[a]	
A135	B150	**C128**
A190	B220	**C160**
Unique subunits		
A14	B12.5	C11
A34.5	B16	C25
A43	B32	*C31*
A49		*C34*
		C37
		C53
		C82

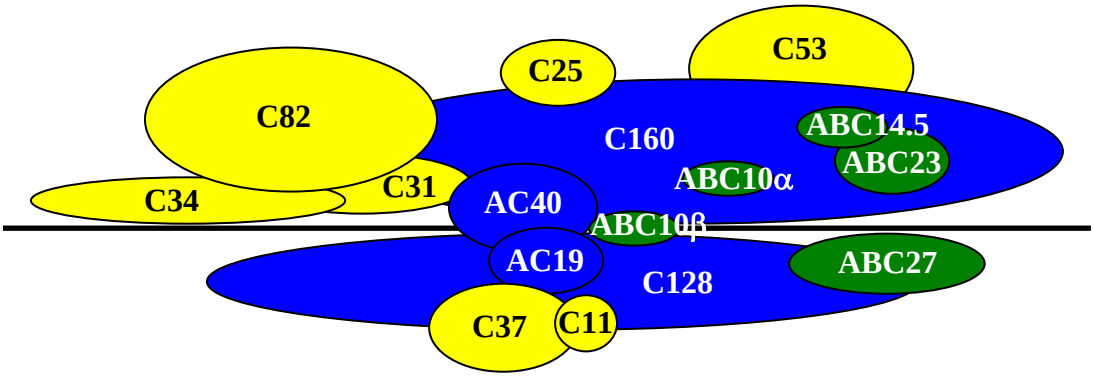

Figure 1.1. **Subunit interactions of the RNA pol III complex.** The DNA is represented as a black line, the core subunits in blue, the shared subunits in green, and the unique subunits in yellow.

complexes that are ready for transcription, have been seen in many organisms, such as yeast (19), rat (20), frog (21), and human (22), and pol III holoenzymes have been seen in humans (23), but no such complex has been seen in yeast. The existence of holoenzymes and the observation that the RNA polymerases share subunits raises the important question of how the polymerases are regulated.

Promoter Structure and Classification

In order to determine how the polymerases are regulated, a discussion of promoter structure and transcription factors is required. Pol II promoters are complex and have various effectors to increase or decrease the transcription of a particular mRNA. These regulators are required to ensure that inappropriate proteins are not unnecessarily expressed, as this would be a waste of cellular resources and would interfere with cell function. Pol I promoters are not as complex as those for pol II, but have similar sequence elements (enhancers, initiators) and are stringently species-specific (24).

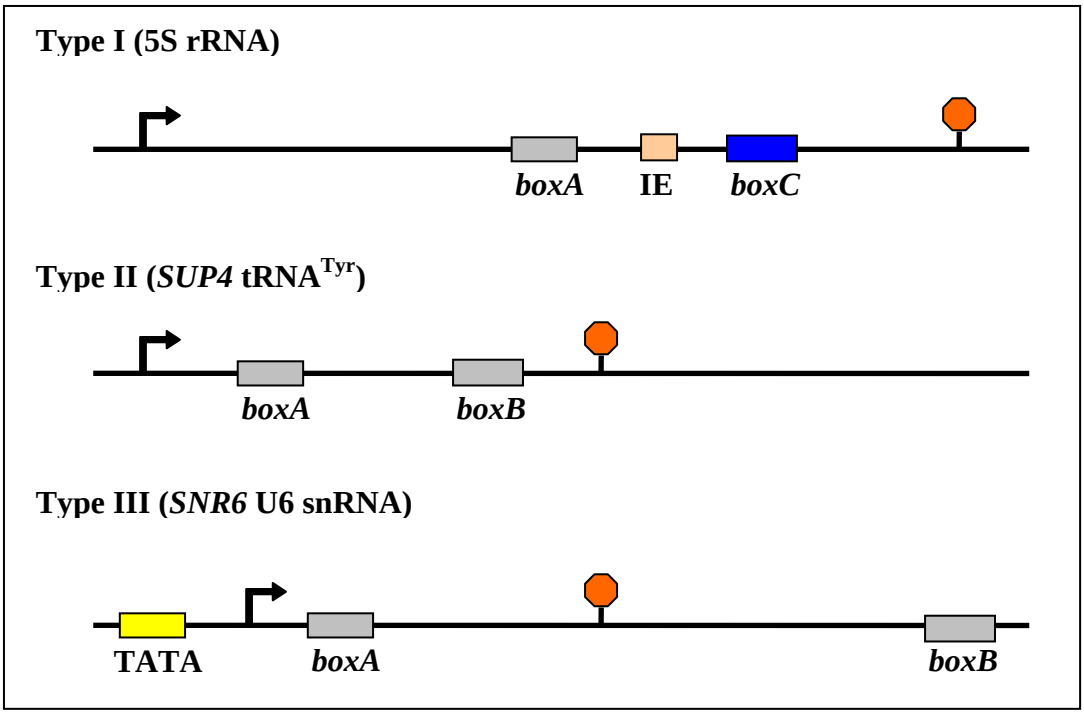

Figure 1.2. **Examples of the three types of RNA pol III promoters.** The DNA is represented as a black line, the transcription start site (+1) as a bent arrow, and the terminator as red octagon. The rectangles indicate important protein binding sites discussed in the text. Drawings not to scale.

The polymerase III system genes can contain one of three types of promoters, called type I, type II, and type III (Figure 1.2). Although these genes all contain gene-internal binding sites for transcription factors, they differ in structure and the types of RNA they produce. Type I promoters are found only in 5S rRNA genes, and consist of *boxA*, *boxC* and an intermediate element. Type II promoters are present in tRNA genes and are the most common form found in pol III genes. These also contain an internal *boxB* region and a *boxA* internal sequence that is homologous to that found in type I promoters. Type III promoters in higher eukaryotes do not contain *boxA* or *boxB* components, though lower eukaryotes, like *Saccharomyces cerevisiae*, do contain functional versions of these elements (25). Genes with type III promoters generally contain a functional TATA box at –30 (relative to the start site of transcription), which is bound by the TATA-binding protein (TBP). TBP is a very important component of all three types of

eukaryotic transcription, which will be discussed following a description of the transcription factors that bind the gene-internal sequences discussed so far.

Isolation and Characterization of Pol III Transcription Factors

The table below shows the number and sizes of the subunits that make up the transcription factors involved in pol III transcription. The role of each of these factors is discussed below.

Table 1.2. **RNA pol III transcription factors of *Saccharomyces cerevisiae* and their subunits.** Subunits are listed by weight in kDa. Named subunits are indicated in parentheses.

TFIIIA	TFIIIB	TFIIIC
50	27 (TBP)	55
	67 (Brf1)	60
	68 (Bdp1)	91
		95
		131
		138

TFIIIA

The first pol III transcription factor purified to homogeneity was TFIIIA, which is involved in transcribing type I promoter genes. TFIIIA purified from yeast (26) is a 50-kDa protein that contains nine zinc-finger domains involved in DNA binding, and its sole essential function in yeast is to bind the promoter of the 5S rRNA gene. To accomplish this, TFIIIA binds the gene-internal *boxC* and acts as a "bridging" factor to place TFIIIC on the *boxA* sequence. TFIIIC then recruits TFIIIB, which recruits pol III to initiate transcription correctly. On genes where *boxA* and *boxB* are both present, TFIIIA is not required, but TFIIIC binds to *boxA* and *boxB* and places TFIIIB on the DNA in the appropriate position to begin transcription.

TFIIIC

TFIIIC, which was first purified from yeast by using a combination of ion exchange chromatography and exploiting its affinity for *boxB* and tDNA sequences, is a large complex consisting of several polypeptides arranged in two globular domains (~300 kDa each, 27) and connected by a flexible linker (28 and Figure 1.3). This flexible region allows TFIIIC, also known as τ, to interact with variably spaced *boxA* and *boxB* elements, some separated by as many as 243 base pairs (25). On genes where these elements are separated by a distance beyond the flexibility of TFIIIC, the DNA is looped out to allow for binding (27).

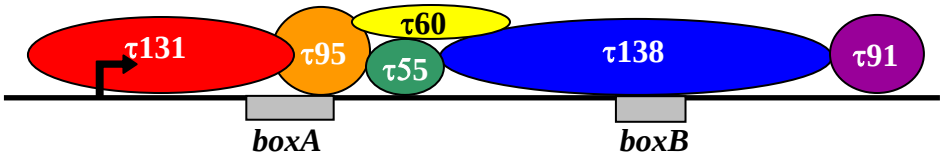

Figure 1.3. **Subunit interactions of the TFIIIC complex on a tRNA gene.** Interactions determined by photochemical crosslinking experiments place the 131-kDa subunit near the transcription start site in a position to interact with TFIIIB (see text and Figure 1.4). The DNA is represented as a black line, and the bent arrow indicates the start site of transcription (+1).Subunit names indicate size in kDa. Drawing not to scale.

The 131-kDa subunit of TFIIIC, or τ131, has been shown by photochemical crosslinking to contact the Brf1 subunit of TFIIIB (29), mutations to *boxA* can change the start site of transcription (30), a yeast U6 gene can be converted from pol II to pol III specificity *in vivo* by addition of TFIIIC (31), and a yeast two-hybrid screen showed that τ131 interacts with the 53-kDa subunit of pol III (32), indicating that TFIIIC is involved in both placement of TFIIIB on the DNA and recruitment of pol III (see Figures 1.1 and 1.4). τ131 has also been shown to be involved in the recruitment of the Bdp1 subunit of TFIIIB, as a deletion of the second tetratricopeptide repeat (TPR) motif caused defective recruitment of Bdp1, which was rescued by overexpression of Bdp1 both *in vivo* and *in vitro* (33).

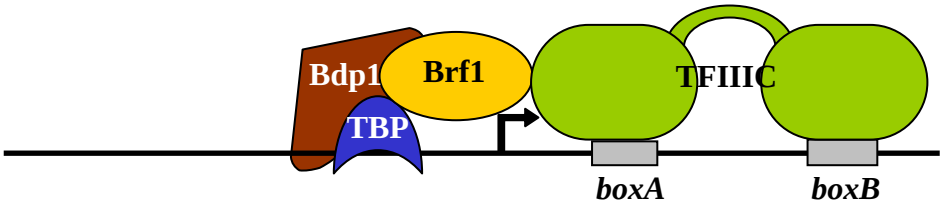

Figure 1.4. **Assembly of the TFIIIB complex *in vivo* via the assembly factor TFIIIC.** The DNA is represented as a black line and the bent arrow indicates the start site of transcription (+1), with TBP ~30 bp upstream of +1. Drawing not to scale.

TFIIIB

The third and most central transcription factor is known as TFIIIB, which is made up of three proteins (29): TBP, the TFIIB-related Brf1, and the pol III-specific Bdp1 (originally referred to as B double prime). Unlike TFIIIA and TFIIIC, there is no evidence of a stable TFIIIB complex in solution without DNA (34). The genes for these proteins have been cloned from yeast: TBP *SPT15* (35), Brf1 *BRF1* (36), Bdp1 *TFC5* (37), and the proteins purified. TFIIIB has been fully reconstituted from yeast *in vitro* using recombinant proteins (37), and was shown to be the transcription initiation factor proper, while TFIIIA and TFIIIC act mainly as assembly factors (38).

Brf1 (Figure 1.5) is a 67 kDa protein that was identified as a suppressor of a temperature-sensitive mutation in yeast TBP and whose N-terminal half has homology to TFIIB, a pol II-specific factor (36). As with almost all of the proteins discussed so far, Brf1 is essential to the

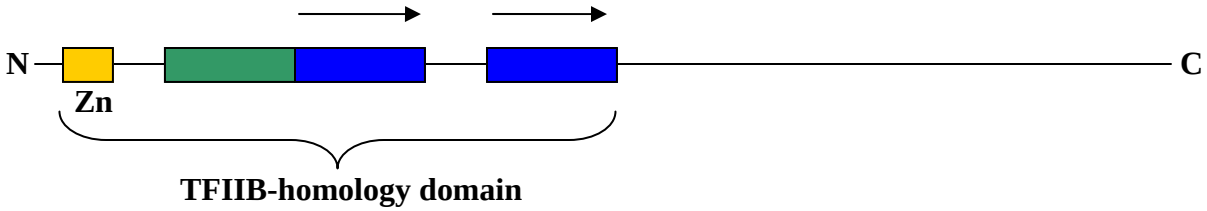

Figure 1.5. **Diagram of the domains of yeast Brf1.** Zn is the zinc-finger domain, and the colored rectangles represent regions that are homologous with yeast TFIIB. The arrows indicate imperfect repeats, and the bracket shows the extent of the N-terminal TFIIB-homology domain. The C-terminal half contains Brf1-specific domains. Drawing not to scale.

growth of yeast, as deletion was strain-lethal (36, 32). Yeast extracts containing Brf1 deletion mutants were tested for transcriptional activity with pol I, II, and III and it was found that pol I and II transcription were unaffected, while pol III transcription was diminished (36). The presence of Brf1 in TFIIIB was established by Western blotting using rabbit antibodies to Brf1 (36). Brf1 was also shown to interact with pol III subunit C17, indicating the role of Brf1 in recruiting pol III to the promoter (39).

Bdp1 is a 68 kDa protein that was identified from yeast extract based on its ability to form active TFIIIB with recombinant TBP and Brf1 (37). Microsequencing of an ~90 kDa band excised from a 10% SDS-PAGE gel allowed for design of degenerate oligonucleotide primers to amplify Bdp1 from genomic DNA, pointing to the gene called *TFC5* (37). To resolve the issue of anomalous migration during SDS-PAGE, *TFC5* mRNA was analyzed using RACE (rapid amplification of cDNA ends) and also cloned into *E. coli* to produce a His_6-tagged Bdp1. The RACE analysis showed no introns and the His_6-tagged Bdp1 migrated exactly as wild type (wt)Bdp1 on SDS-PAGE gels (37). *TFC5* was also shown to be essential to yeast cell growth, as disruption of the gene sequence by insertion of the *URA3* gene was lethal (37). Most importantly, DNaseI footprinting showed that recombinant TFIIIB was indistinguishable from wtTFIIIB on the $tRNA^{Tyr}$ (*SUP4*) and U6 snRNA (*SNR6*) genes, and a truncated form of Bdp1 (comprised of amino acids 40-487) is still functional for transcription (37).

Thorough investigation of the contacts in the TFIIIB-DNA complex has been undertaken by several groups using techniques such as hydroxyl radical (40-42) and DNaseI footprinting (41, 43), photochemical crosslinking (44, 45, 42), mutagenesis of DNA/proteins (40, 41, 43), complementation (42), and deletion (46, 43). These will be addressed after a discussion of the third protein in the TFIIIB complex, TBP.

TBP was first isolated as part of the TFIID complex from the yeast pol II system (47). After TATA-like sequences were seen in pol III promoters, the question of TBP as a pol II/III factor needed to be addressed. Several genetic and biochemical experiments showed that the TFIID fraction was required for transcription of pol II genes (reviewed in 48), as well as the *in vitro* transcription of the U6 (pol III) genes (49), and temperature-sensitive TBP mutants showed a loss of transcription for all three polymerases (50), which was confirmed by placing TBP under the control of a copper-inducible promoter (50). Schultz *et al.* (51) also showed that TBP can recover specific initiation by pol I, II, and III in yeast extracts where TBP was lacking.

The *SPT15* gene, which encodes TBP, was isolated via mutagenic analysis of repressors of Ty retrotransposon integration (52), which is involved with the pol III transcription system (53). Microsequencing of a 25 kDa band from an SDS-PAGE gel allowed for the development of degenerate oligonucleotide primers to amplify the TBP gene from yeast genomic DNA, and deletion of this gene was also shown to be lethal in yeast (54). TBP is a highly basic, 240-amino acid, 27-kDa protein (54), with a conserved 180 amino acid C-terminal domain consisting of two repeated sequences that are 28% identical (55). The N-terminal domain is not phylogenetically conserved (reviewed in 51, 56, and 57), but has been implicated in regulation of TBP binding by oligomerization (58-64), though this assertion is controversial, as discussed below.

The amino acid sequence of TBP contains two subdomain repeats separated by a basic region that are essentially identical in structure, with each repeat containing five β-strands and two α-helices. The β-strands form a concave, 10-strand, antiparallel β-sheet that is bracketed by two perpendicular, short α-helices and paralleled by two long α-helices (Figure 1.6). The S2 and S2' β-strands contain the loops where the kink-forming phenylalanine residues reside (57). The repeated domains of the C-terminal TBP domain (also called the core or TBPc) are involved in

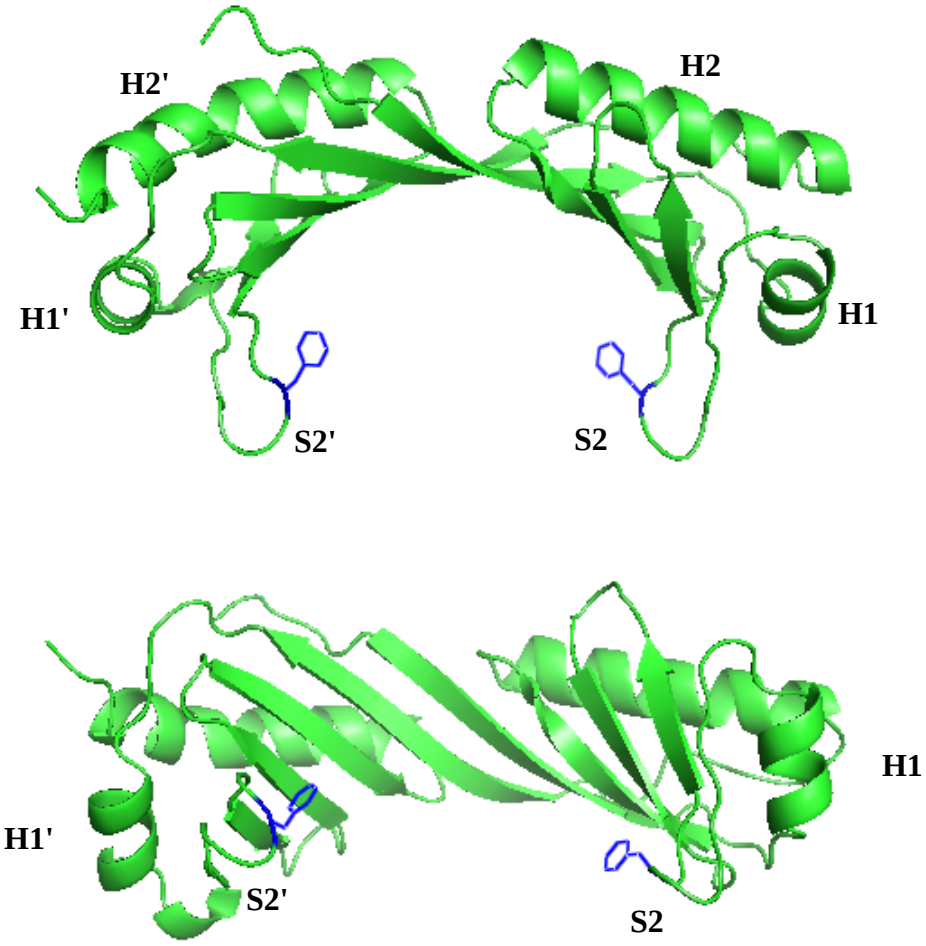

Figure 1.6. **Crystal structure of yeast TBP.** Helices and sheets discussed in the text are labeled. The intercalating phenylalanine residues are shown in blue. The bottom panel shows the view from underneath the saddle, looking up at the DNA-binding interface. This figure was generated using PyMol and the 1YTB file from the Protein Data Bank.

binding and bending the TATA box DNA (Figure 1.7). This is accomplished by the intercalation of two phenylalanine residues between bases 1 and 2 and bases 7 and 8 of the 8-bp TATA box (55). These kinks in the DNA backbone introduce a bend on the order of ~80° and cause an unwinding of the DNA, which is compensated by supercoiling of the helix, allowing for the maintenance of Watson-Crick base-pairing even in this non-B-form DNA. The minor groove of the DNA binds to the underside of the saddle-shaped TBP and interacts via hydrophobic

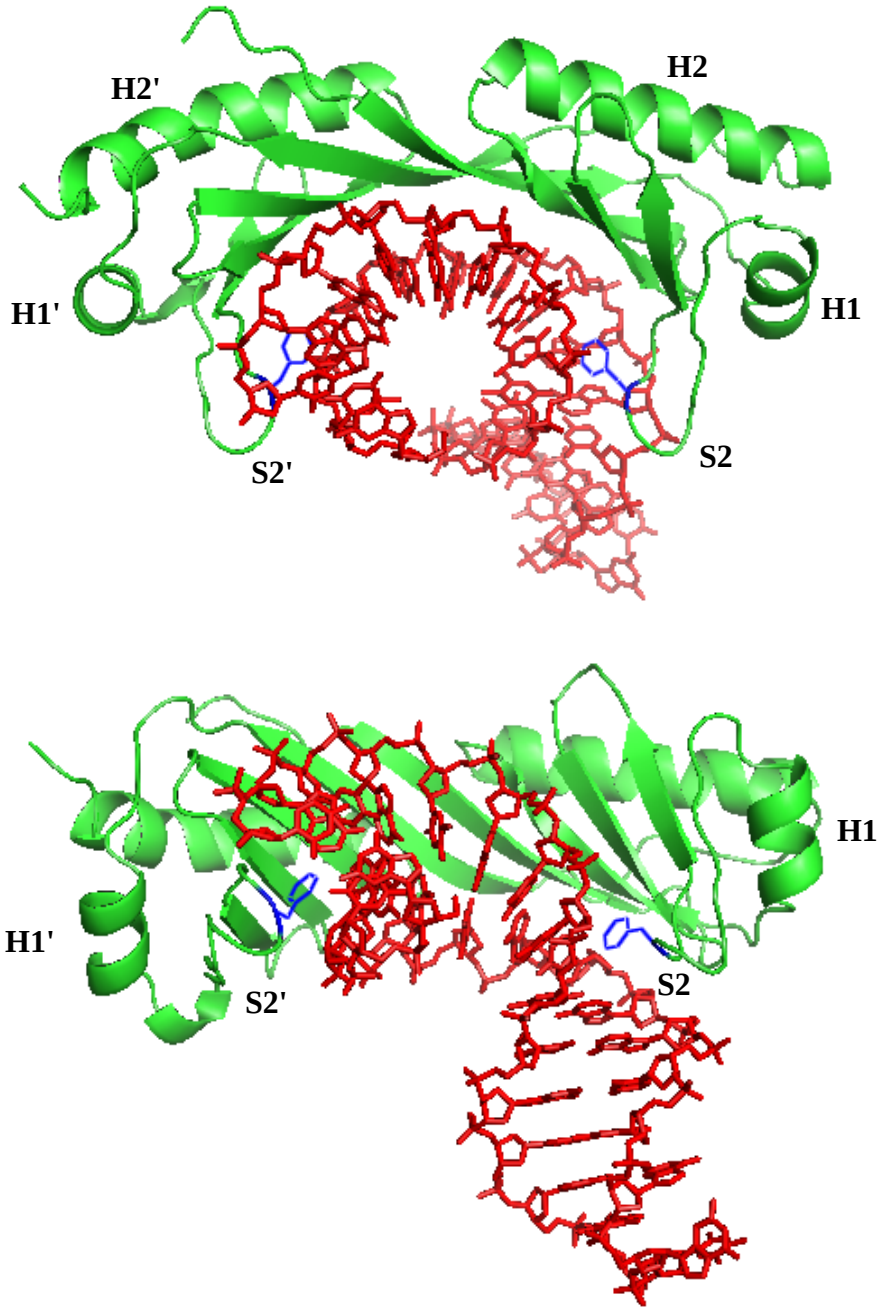

Figure 1.7. **Crystal structure of yeast TBP bound to DNA.** TBP is shown in green and the DNA in red. Helices and sheets discussed in the text are labeled. The intercalating phenylalanine residues are shown in blue. In the bottom panel, the bend induced by TBP in the DNA is clearly visible. This figure was generated using PyMol and the 1YTB file from the Protein Data Bank.

interface with the minor groove base edges and sugars, burying about 3100Å^2 of surface area on TBP binding (55).

The C-terminal end of TBP was seen to contact the 5' end of the TATA box, even though the protein is almost symmetrical. Based on the co-crystal structure of yeast TBPc (yTBPc) with the TATA box from the *CYC1* promoter, the relatively higher positive charge on the surface of TBP that binds the 3' half of the TATA box may help in deforming the less-flexible TAAA sequence and account for the preference for binding observed (55). Also, three side chains (Leu163, Val119 and Val80) in the crystal structure of *Arabidopsis thaliana* TBP2 with AdMLP TATA DNA were seen to make van der Waals contacts with three adenines (-30, -28, -26), specifying rightward binding by allowing TBP to distinguish A from T at the 5' end of the asymmetrical AdMLP TATA box (TATAAAAG, 65). Experiments performed with TATA DNA made more flexible by substitution of hydroxymethyluracil (hmU) for T at the sites of the kinks also indicate that, on its own, TBP will equilibrate to the rightward orientation seen in the crystal structures (66). Kinetic studies using Förster resonance energy transfer (FRET) (67) also support a slight preference (60/40) for rightward orientation of TBP on TATA DNA in solution. It should be noted, however, that TBP interactions with other proteins found in transcription factors change this ratio, as in the yeast U6 gene, where TFIIIC places TFIIIB on the DNA only in the rightward direction (68), and in the pol II system, where TFIIB causes TBP to assume a rightward position 80% of the time on the AdMLP promoter (69).

TBP Dimerization: Relevance *In Vivo* and Function of the N-Terminal Tail

Crystal structures of yTBPc showed two molecules of protein in a dimer configuration (57), which led several labs to investigate whether such dimerization could occur *in vivo* and might be involved in regulation of transcription by preventing TBP binding to DNA. The

intrinsic fluorescence of tryptophan 26 (W26) in the N-terminus of yTBP was used to show that the full length protein assumes oligomeric states (tetramer and octamer) in solution at nanomolar concentrations, and that binding of TBP to DNA showed a large red shift of W26 emission, from 310-353 nm (59). This is consistent with total unfolding of TBP when compared to spectra caused by guanidine hydrochloride and thermal denaturation, indicating at the very least a change in the environment of W26, if not a complete unfolding of the N-terminal tail from the body of the protein (59). Upon proteolytic cleavage of TBP to remove residues 1-40, the shift of W26 is identical to that observed upon DNA binding, so the interaction of the N-terminal tail of TBP may be with the carboxy termini of other molecules of TBP in solution (59).

The observed oligomerization does not seem to affect the kinetics of binding, since the values were the same for a range of TBP concentrations (20-600 nM), and the observed on-rate value of $1.66 \times 10^5 M^{-1}s^{-1}$ (59) was almost the same as the value of $2.6 \times 10^5 M^{-1}s^{-1}$ reported by Hoopes *et al.* (70) using different techniques, though this could merely be indicative of the fact that TBP does not dimerize at these concentrations. Kinetic studies done with an N-terminal deletion of TBP showed a slower rate for association with DNA, indicating that the N-terminus is involved in increasing the rate of TBP-DNA interaction (59).

Gel filtration was used to show dimerization occurring with human TBP at 100 nM, but no higher order oligomers were observed at this concentration (58). Crosslinking by 1,6-bismaleimidohexane (BMH) also showed dimers, but no evidence of higher order oligomers, and dimers were competed away by addition of TATA DNA (58). Jackson-Fisher *et al.* (60) showed dimerization of yeast TBP using BMH at 2 μM and also showed heterodimerization of yTBP with h180C (the core of human TBP). They also claimed that the work of Petri *et al.* (71) showing no dimerization was the result of TBP inactivation that occurred during preincubation in

the absence of DNA at 30°C, determined by performing experiments under the same conditions. However, Jackson-Fisher *et al.* failed to include 1 mM $CaCl_2$ in their reaction conditions while mimicking those of Petri *et al.* The effect of the absence of this extra salt may have something to do with the discrepancy, as well as the increased sensitivity of fluorescence measurements compared with determinations of kinetic parameters by electrophoretic mobility shift assay (EMSA), where TBP-DNA complexes are subject to dissociation by electrophoresis (70).

Analytical ultracentrifugation showed that yTBP exists in solution in a monomer-tetramer-octamer equilibrium, with the majority of the molecules in the octamer state at the micromolar concentrations used (62). Models that included dimers did not fit the data as well as models that excluded dimer parameters, indicating that dimeric TBP is not predominant at these concentrations and under these conditions (62). Sedimentation equilibrium studies also showed no evidence of dimerization of full-length yTBP at 30°C at physiological concentrations (72), though buffer conditions used (Campbell's buffer contained 10% glycerol) were different from those of Daugherty *et al.* (62). Using fluorescence, a more rigorous and quantitative method for determining dimerization than GST pull-down assays or crosslinking used by others (58), to investigate TBP interactions over a wide range of KCl (58 mM-1 M) and temperatures (4°-37°C), Daugherty *et al.* (63) confirmed this lack of dimers and asserted a predominance of octamers, while showing that lack of glycerol and $MgCl_2$ in buffers does not affect the distribution of TBP oligomers.

The N-terminal region of TBP contains a region of ten amino acids just N-terminal to the core that are required for viability of two TBP mutants that are deficient in TFIIA and Brf1 interactions (R137A and A140R, respectively) (61). Removal of the N-terminus restores activated transcription to DNA-binding mutants of TBP, indicating that the N-terminus inhibits

transcription *in vivo* (61). Because the N-terminus acts outside of the DNA-binding region to activate transcription, this may account for its lack of phylogenetic conservation (55, 65), as it evolved to interact with factors specific to each organism (61).

While TBP is known to be present in yeast nuclei at a concentration of 6.3 μM (referenced in 62), there is no information on how it is distributed in the nucleus (62), and it is likely to be found complexed with other transcription factors (72) or on DNA, as free TBP will bind DNA non-specifically and can nucleate transcription in an incorrect location (58). A speculative function of self-associated TBP may be to act as a reservoir to facilitate rapid changes among polymerase I, II, and III activities and may function as a way to keep TBP in the nucleus, as it could diffuse out passively through the nuclear pore complex (63).

Attempts to elucidate the *in vivo* nature of TBP dimers were the impetus for mutagenesis of residues in the concave underside of TBP, which contains the DNA-binding domain and the dimer interface region observed in the crystal structure (55, 65). Mutation of arginine 171 to glutamate was shown to reverse derepression of transcription typical of N69R (a dimer-destabilizing mutant), apparently by stabilizing the dimer population (64). This was not due to disruption of TBP structure or function, as R171E was able to support viability on its own after wtTBP was shuffled out using 5-fluoroorotic acid (5-FOA) selection, although the growth was slower. Brf1 disrupts TBP dimerization, as visualized by less crosslinking of TBP upon addition of increasing amounts of Brf1, while depletion of Brf1 with antibodies allowed for crosslinking of TBP to be restored (73). Brf1 (1-262) and Brf1 (263-596) also could not inhibit TBP dimerization (73), even though Brf1 (263-596) has affinity for TBP (74), and others (42) have shown that combination of these derivatives can substitute for full-length Brf1. GST pull-down assays showed that Brf1 binds TBP dimers and causes them to dissociate, possibly by interaction

of the N-terminal TFIIB-homology domain with the C-terminal stirrup of TBP (73). Though much work has been done on TBP dimerization, for experiments performed under our conditions (66), it does not appear to be an issue.

TBP-Induced Bend Angle of Promoter DNA: Magnitude and *In Vivo* Relevance

The co-crystal structure of TBP and the TATA box showed a severe bend in the DNA, caused by two kinks between base pairs at either end of the TATA box where phenylalanine residues intercalate into the DNA (55). The structure also showed unwinding of the helix and flattening of the minor groove in the region between the kinks bound to the underside of TBP (55). This issue is also slightly controversial, with some groups arguing that the bend angle observed is an artifact of the osmolytes used in crystallization (75), that the TATA box inherently assumes a bent conformation in the absence of TBP (76, 77), whether DNA bending occurs before or after binding of TBP (78, 70), if the observed non-diffusion-limited kinetics are due to an initial recognition step followed by bending (79), or if the two processes are even separate events (80).

TBP interacts with the minor groove of the TATA box (Figure 1.7 and 81), and gel retardation assays show that TBP induces a bend in the DNA upon binding (78). EMSA can be used both to determine the magnitude of the bend and to derive kinetic parameters for TBP-DNA complex formation (70). The speculated function of the bend *in vivo* is to bring the components of the transcription machinery closer together in order to allow interactions between downstream and upstream DNA elements (Figure 1.8 and 78). This seemed to be borne out by further work from the Hawley lab (82) indicating that, in the pol II system, increased bend angles correlated with increased complex stability and increased levels of transcription. A circular permutation assay was used to determine that TATA boxes of different sequence produce different TBP-

induced bend angles, and that complexes with longer lifetimes contained DNA that was bent more than that found in complexes with shorter lifetimes (Table 1.3 and 82). The sequence of the TATA box for two pol II promoters affects the kinetics of binding (71) and bending can explain the observed slow kinetics of TBP binding without invoking TBP dimerization. FRET studies indicated that two intermediate species are formed during TBP-DNA interactions, both with DNA bent similarly to the final complex (80). This work also confirmed the on-rates obtained by other groups (59, 70, 71, 83-85), and the second-order nature of TBP-DNA binding kinetics, as well as the fact that binding and bending are simultaneous, as even the first intermediate described contains bent DNA (80).

Table 1.3. **TBP induces bends to different degrees in DNA of varying sequence.** In the RNA pol II system, TBP-induced bending of promoter DNA is positively correlated with increased transcription and complex lifetime. Adapted from reference 82.

Mutant	Sequence	Bend Angle (°)	Lifetime (min.)	Relative Transcription
27T	TATATAA	106	185	0.73
WT	TATAAAA	93	90-100	1.00
28T27T	TATTTAA	93	70	0.72
30T	TTTAAAA	87	40	0.68
28T	TATTAAA	80	7-9	0.35
31C	CATATAA	63	4	0.41
29A27T	TAAATAA	59	1.5	0.15
25C	TATAAAC	34	8	0.24
29A	TAAAAAA	<34	1	0.11

Since TBP is also involved in transcription by pol III, and the pols share the characteristics discussed above, it is possible that the bend angle induced by TBP binding to promoter DNA may have a similar effect on transcription. Though increased bend angle correlates with increased levels of transcription, no positive correlation between affinity and complex lifetime was observed by Wu *et al.* (75), as seen by others (82). Also, it should be noted that formation of the TFIIIB-DNA complex introduces another bend downstream of the TATA

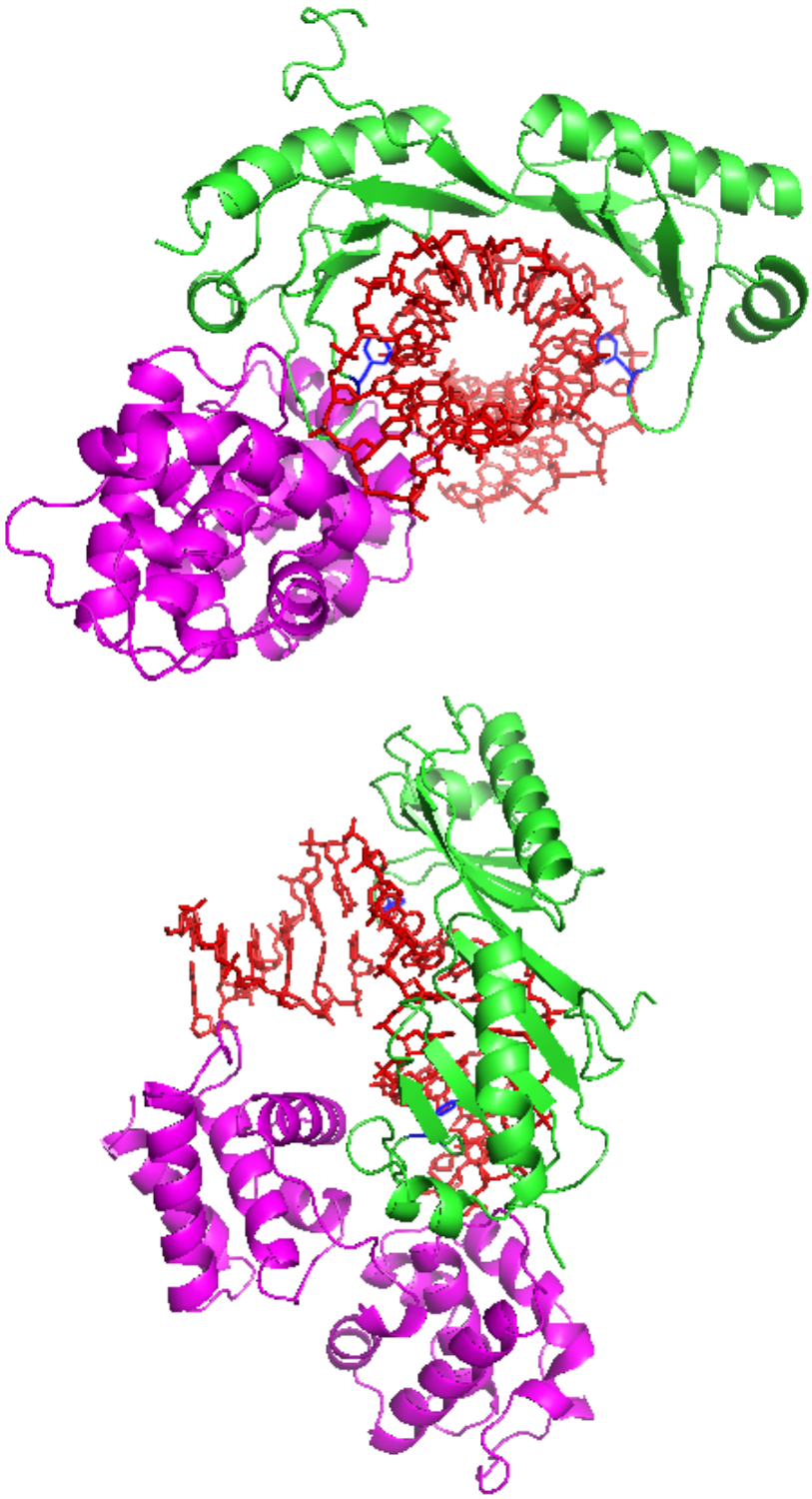

Figure 1.8. **Crystal structure of yeast TBP and TFIIB bound to DNA.** TBP is shown in green, TFIIB in purple, and the DNA in red. The intercalating phenylalanine residues are shown in blue. The bottom panel clearly shows how the bend induced by TBP in the DNA allows the DNA to interact with both proteins. This figure was generated using PyMol and the 1VOL file from the Protein Data Bank.

box and upstream of the start site of transcription (86). The functions of TFIIIB will be discussed after an overview of the protein-DNA contacts found in this complex.

Properties of the TFIIIB-DNA Complex

The role of TFIIIC as a required factor for the proper placement of the TFIIIB complex on the promoter of genes containing type I and II promoters (38) was used as the starting point for studies to determine the assembly order of TFIIIB by TFIIIC on the *SUP4* gene (29). A weak Brf1-TFIIIC-DNA complex was stabilized by the entry of TBP, as evidenced by increased DNaseI protection of the *SUP4* gene and increased levels of photochemical crosslinking between Brf1 and the τ131 subunit of TFIIIC (29). Addition of Bdp1 does not occur until TBP is present, as only then do the footprints and heparin resistance of the complexes match (29). TBP only weakly crosslinks to upstream DNA and does not act in a DNA-binding capacity on *SUP4* promoters (as they lack a TATA box), but causes a conformational change in Brf1 that reveals a hidden DNA-binding domain (29). Entry of Bdp1 is speculated to cause a similar conformational change that gives the TFIIIB-DNA complex its “extreme stability” (87).

This order of addition was confirmed using DNaseI and hydroxyl-radical footprinting, as well as a deletion analysis of Bdp1 (41). The robust nature of the TFIIIB-DNA complex was evidenced by the fact that, for *SNR6* transcription, no single amino acid of Bdp1 was found to be essential, indicating that Brf1 or TBP was able to compensate for the missing residues (41). Even Bdp1 that was severely truncated was still functional for *SNR6* transcription *in vitro* (41). It was only when these deletions were combined that TFIIIB complexes became inactive for *SNR6* transcription (41). The smallest Bdp1 required for *SNR6* transcription was a 176-amino acid region (from 224-400), and even deletions made to this region in the full-length protein were still active for transcription (41).

The same was true of TFIIIC-dependent transcription of the *SUP4* gene: although it also required some residues N-terminal and C-terminal to the core region and was stable to N- and C-terminal truncations, combinations of these deletions were inactive on the *SUP4* gene (41). Bdp1 mutants that were able to form TFIIIB complexes but unable to transcribe DNA were isolated (224-378 and 241-378), and further analysis showed that two regions of Bdp1 (378-418 and 186-224) are involved in interactions with pol III (41). DNaseI footprinting performed with mutant Bdp1 showed that N- and C-terminal truncations generated footprint changes at the downstream end of the TATA box; therefore, these two domains must be near each other in the TFIIIB-DNA complex (41). Since addition of Bdp1 causes changes in the TBP-Brf1-TFIIIC-DNA complex footprint and gives the TFIIIB-DNA complex considerable heparin resistance, Bdp1 appears to act as a scaffold for TFIIIB complex formation and clamps down on DNA via protein-protein interactions (41). Bdp1 mutants that are inactive for transcription (Δ272-292 and Δ409-421) have deficient mechanisms for displacing τ120 from the start site of the *SUP4* gene, and are thus unable to recruit pol III for transcription (41).

Much work has been performed to demonstrate the functional redundance between the Brf1 and Bdp1 proteins in the TFIIIB-DNA complex. Brf1 lacking the N-terminal TFIIB-homology domain (Figure 1.5) was still able to nucleate assembly of TFIIIB on *SNR6* DNA when full-length Bdp1 was used (46); however, when Brf1(165-596) was combined with Bdp1 deletion mutants sufficient for transcription with full-length Brf1, transcription defects emerged and implicated a region of Bdp1 (amino acids 355-372) in interaction with pol III, since TFIIIB-DNA complexes stable to electrophoresis but unable to generate transcript were formed (46). The N- and C-terminal halves of Brf1 can complement each other to form a TFIIIB complex on TA-30, a modified *SUP4* gene with a 6-bp TATA box inserted at -30 and the surrounding

sequence made G+C-rich, that was indistinguishable from that formed by wtBrf1 on the same DNA, and stable to both electrophoresis and heparin challenge (42). TBP mutants that prevented binding of wtBrf1 also prevented the C-terminal half of Brf1 from binding TBP (88), and a TBP mutant that was defective for TFIIIB complex formation was found to prevent crosslinking of the N-terminus of Brf1 to upstream DNA on the *SNR6* gene (42). These results show that the C-terminal repeat of TBP interacts with the N-terminal half of Brf1, and the N-terminal repeat of TBP interacts with the C-terminal portion of Brf1, supplying most of the affinity that Brf1 has for TBP (42). The N-terminal portion of Brf1 is responsible for the pol III-recruitment activity of TFIIIB, as shown by a nearly complete loss of TFIIIC-independent transcription on *SNR6* when Brf1 N-terminal deletions were employed (42).

Brf1 requires 15 bp of DNA downstream of the TATA box and only 1 bp upstream to form a stable complex, while TFIIIB requires 15 bp upstream and 15 bp downstream of the TATA box for stable complex formation (40). This lends credence to the “protein clamp” model suggested for the extreme stability of TFIIIB on DNA and gives further insight into how the proteins and DNA in this complex interact. Support for the clamp model also comes from photochemical crosslinking experiments showing extensive contacts between Brf1 and Bdp1 along the entire TFIIIB-DNA complex, as well as crosslinking of the N-terminal portion of Bdp1 to DNA upstream of the TATA box and the C-terminal portion to the DNA downstream of the TATA box (44). Extensive photochemical crosslinking performed on the TFIIIB-*SUP4* complex showed that TBP crosslinks to a TATA-like region at -30, but this requires placement by TFIIIC (45). Brf1 crosslinks within and downstream of the TATA-like region, also in a TFIIIC-dependent fashion, while Bdp1 crosslinks upstream of and near the TATA-like region (45). The C34 subunit of pol III crosslinks downstream of the TATA-like region (at -17 and -16), and

C128 and τ120 mapped to a region (-22, -21, -17, and -12) several base pairs downstream of the TATA-like region (45). Entry of Bdp1 to the complex changes contacts between TBP and DNA, as well as between Brf1 and DNA, and displaces τ120 from the DNA downstream of the TATA-like region (-38 to -23), as shown by photochemical crosslinking (45). Crosslinking also supports the protein clamp structure of TFIIIB on DNA, as Brf1 was seen to map to one side of the complex, while Bdp1 maps to the other side (45). Interestingly, a difference between the TBP-DNA and TFIIIB-DNA complexes was highlighted by crosslinking, showing that addition of Bdp1 causes TBP to crosslink with the major groove at -23, *i.e.*, the downstream end of the TATA box, indicating that Bdp1 may further bend the DNA in this region (45).

This notion of a conformational change upon entry of Bdp1 was confirmed by missing nucleoside selection on the modified *SUP4* gene, TA-30, which showed that addition of Brf1 and Bdp1 abrogated preference exhibited by TBP for flexibility (caused by the missing nucleoside) at position -23, though there was still preference for missing nucleosides in a region at the downstream end of the TATA box (86). Placement of 4-nt loops downstream of the TATA box (at positions -8 and -17 or -5 and -15) increased formation of the TFIIIB complex, localizing the Bdp1-induced bend to this region, centered at -15 (86). Since subunits of pol III map to this region using photochemical crosslinking (45), it is possible that this deformation of the DNA is involved in recruitment of polymerase to initiate transcription.

DNaseI footprinting on the *SNR6* gene was used to determine the free energy of formation (ΔG°) for TBP-DNA, B'-DNA, and TFIIIB-DNA complexes and showed that the B'-DNA and TFIIIB-DNA complexes have the same ΔG° but different lifetimes in solution (89). The binding curves determined by this method also show that Brf1 binds cooperatively (n_H = 1.3 ± 0.3) to TBP-DNA and increases the affinity of TBP for the TATA box, since increasing Brf1

allowed for complex formation at lower concentrations of TBP (89). Taking these two factors into consideration, the protein clamp model for TFIIIB-DNA complex formation, complete with extensive protein-protein contacts discussed so far, seems very reasonable. However, since TBP is required for transcription of all three types of pol III genes--including those that lack a TATA box (30)--and since the bends induced by TBP and Bdp1 increase the stability of TFIIIB-DNA (66, 86) regardless of TATA sequence (90), how can the formation of B'-DNA and TFIIIB-DNA have the same $\Delta G°$?

If the DNA is kinetically trapped by the protein clamp of TFIIIB (89), the sequence of the upstream region may have little influence, since TFIIIB-DNA formation is rapid compared to TBP-DNA formation (70) and addition of Brf1 and Bdp1 can trap a population of TBP-DNA complexes in a non-equilibrium distribution (86). The co-crystal structure of Brf1(437-596) with yTBP(61-240) and part of the U6 promoter (-31 to -13) DNA shows that the C-terminus of Brf1 binds to the top of the TBP saddle (91), providing a mechanism for Brf1 to place TBP on the promoter DNA near -30 even when no TATA box is present (30). So, the “lost” energy between the B'-DNA and TFIIIB-DNA complexes might be used to bend DNA further upon addition of Bdp1, may be due to lost TBP-DNA contacts upon addition of Bdp1 (45), or may increase the energy required to dissociate TFIIIB-DNA (89). These biochemical analyses demonstrate a fairly complete picture of contacts found in the TFIIIB-DNA complex and the process of TFIIIB complex formation on the promoter; however, as Cloutier *et al.* (89) note, the lifetime of the TFIIIB-DNA complex ($t_{1/2}$ = 95 minutes) is comparable to the doubling time of yeast in rich medium (90-100 minutes), which leads to the question of how TFIIIB performs its function in transcription and how transcription is terminated, or reinitiation prevented, *in vivo*.

Function of TFIIIB in Promoter Opening and Reinitiation

After TFIIIB assembles (with or without the help of TFIIIC) on the promoter DNA to recruit pol III, the DNA in the promoter region unzips, and after a few rounds of abortive initiation (92), pol III clears the promoter and rapidly transcribes the gene to the terminator, where a poly-T sequence is sufficient to stop transcription by causing a long pause (93). After termination, pol III can be brought back to the still-present TFIIIB complex to begin transcription again, possibly through a mechanism involving looping of the DNA back on itself (94). Reinitiation is much faster and pol III remains committed to the original gene, even when a competing gene is located nearby (93). The terminator sequence is also required for facilitated recycling of pol III to occur, as shown by runoff transcription experiments where slower reinitation was observed (93). TFIIIC-pol III interactions may be responsible for bringing the polymerase back to +1, as τ131 has been shown to interact with ABC10α and C53 (see Figures 1.1 and 1.3) (33, 95). Since TFIIIC can easily be removed by transcribing pol III (96), this recruitment activity may be accomplished via the contacts between TFIIIC and TFIIIB described earlier, which would keep TFIIIC nearby after displacement by pol III and facilitate reinitiation.

The role of TFIIIB in pol III recruitment has already been discussed above, but several laboratories have done work indicating that TFIIIB may also have a role in addition to its pol III-recruitment function. Photochemical crosslinking studies showed that formation of the TFIIIB complex removes TFIIIC from the transcription start site (87). Potassium permanganate ($KMnO_4$) footprinting, which exploits the sensitivity of unpaired T bases to oxidation by permanganate, showed that melting of the promoter is temperature-dependent and reversible, with two parts of the transcription bubble opening independently (97). Crosslinking of pol III subunits on SUP4 DNA was seen to require box B and TFIIIB, as was specific placement of pol

III (98). TFIIIB assembled on *SNR6* genes using either mutant Brf1 or Bdp1 was unable to initiate transcription on linear DNA, but was able to transcribe supercoiled DNA or linear DNA made more flexible by the replacement of T with hmU (43). These “closed” complexes were able to be restored by unpairing 3- or 5-bp stretches of DNA in the transcription bubble observed in “open” complexes (99). More specifically, Brf1 defects were rescued by opening parts of the sequence downstream of +1 (+2 to +6) and Bdp1 defects were rescued by opening parts of the sequence upstream of +1 (-9 to -5) (99). TFIIIB, therefore, is involved in two steps of promoter opening: melting of the upstream promoter to open the transcription bubble (via Bdp1) and propagation of the melted bubble downstream (via Brf1) (99). TFIIIB is also involved in processing pol III transcripts, as *RPR1* (the gene for the RNA component of RNase P) was able to suppress the temperature-sensitive phenotype of a strain of yeast bearing Bdp1 deleted between 253 and 269 (100). RNase P is responsible for maturation of tRNAs, and the Bdp1 deletion strain was also deficient in tRNA maturation (100), indicating a post-transcription function for TFIIIB.

Experiments performed on the U6 gene showed that single strand breaks in the transcribed strand of DNA within the region upstream of +1 rescued the defect in promoter opening observed with mutant Brf1 and single strand breaks in the region from -9 to -5 in the transcribed strand rescued the mutant Bdp1 defect that prevents downstream propagation of the transcription bubble, indicating a polarity imposed by TFIIIB in promoter opening (101). Photochemical crosslinking experiments performed on U6 DNA with TFIIIB and pol III showed that though Brf1, Bdp1 and several pol III subunits (C160, C128, C82, C34) are all present at -8 and -7, at the upstream end of the transcription bubble observed in open complexes, TFIIIB is only indirectly involved in promoter opening (102). Evidence for this comes from crosslinking

data showing that the region of Bdp1 that is near the transcription bubble (425-485) is not the region (355-372) that when deleted causes defects in transcription complex formation (102). Finally, TFIIIB was shown to be sufficient for reinitiation on short class III pol III genes, while longer genes required TFIIIC to generate an equal amount of reinitiation (103).

The observation that Brf1 and Bdp1 abrogate the preference of TBP for missing nucleosides at the downstream end of the TATA box (86), along with the changes in photochemical crosslinking patterns caused upon binding of Brf1 and Bdp1 (45), indicate that entry of these proteins into the TFIIIB-DNA complex causes changes in the TBP-DNA complex at the downstream end of the TATA box. The sequence preference of TBP for the TATA box has already been established (TATAa/tAa/t, 104), but TBP can bind to the TATA box in either orientation (see above). One goal of this work is to determine the effect of entry of Brf1 and Bdp1 to the TFIIIB complex on TBP-DNA contacts at the downstream end of the TATA box. In order to attain unidirectional binding, and thus allow determination of sequence preference for TBP in the absence or presence of Brf1 and Bdp1, a mutant TBP that binds the sequence TGTA as well as TATA (TBPm3, 105), was used to perform an iterative *in vitro* selection on a probe with the last four bases of the TATA box randomized. This selection was repeated with the entire TFIIIB complex, and consensus sequences determined.

The second question to be addressed in this work is the effect of the TBP-induced bend angle on the bend angle within the TFIIIB-DNA complex and on transcription. As shown in Table 1.3, increased bending in the pol II system correlates positively with increased rates of transcription (82). Since these polymerases share several subunits (Table 1.1), and both transcription systems use TBP, it is possible that this correlation might also be observed in the pol III system. TBP- and TFIIIB-induced bend angles were determined using a circular

permutation assay, and *in vitro* transcription was performed to ascertain the effect of DNA bending on transcription.

References

1. Engelke, D. R., Shastry, B. S. & Roeder, R. G. (1983). *J. Biol. Chem.* **258,** 1921-31.
2. Roeder, R. G. (1983). *J. Biol. Chem.* **258,** 1932-41.
3. Jaehning, J. A., Woods, P. S. & Roeder, R. G. (1977). *J. Biol. Chem.* **252,** 8762-71.
4. Wang, Z., Luo, T. & Roeder, R. G. (1997). *Genes Dev.* **11,** 2371-82.
5. Sklar, V. E., Schwartz, L. B. & Roeder, R. G. (1975). *Proc. Natl. Acad. Sci. USA* **72,** 348-52.
6. Sklar, V. E. & Roeder, R. G. (1976). *J. Biol. Chem.* **251,** 1064-73.
7. Adman, R., Schultz, L. D. & Hall, B. D. (1972). *Proc. Natl. Acad. Sci. USA* **69,** 1702-6.
8. Hager, G. L., Holland, M. J. & Rutter, W. J. (1977). *Biochemistry* **16,** 1-8.
9. Wandzilak, T. M. & Benson, R. W. (1978). *Biochemistry* **17,** 426-31.
10. Hammond, C. I. & Holland, M. J. (1983). *J. Biol. Chem.* **258,** 3230-41.
11. Roeder, R. G. & Rutter, W. J. (1969). *Nature* **224,** 234-7.
12. Roeder, R.G. (1969). Ph.D. Thesis, University of Washington, Seattle, WA.
13. Weinmann, R. & Roeder, R. G. (1974). *Proc. Natl. Acad. Sci. USA* **71,** 1790-4.
14. Valenzuela, P., Hager, G. L., Weinberg, F. & Rutter, W. J. (1976). *Proc. Natl. Acad. Sci. USA* **73,** 1024-8.
15. Darst, S. A., Kubalek, E. W. & Kornberg, R. D. (1989). *Nature* **340,** 730-2.
16. Gabrielsen, O. S. & Sentenac, A. (1991). *Trends Biochem. Sci.* **16,** 412-6.
17. Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H. & Severinov, K. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 892-7.
18. Chiannilkulchai, N., Stalder, R., Riva, M., Carles, C., Werner, M. & Sentenac, A. (1992). *Mol. Cell. Biol.* **12,** 4433-40.
19. Koleske, A. J. & Young, R. A. (1994). *Nature* **368,** 466-9.

20. Ossipow, V., Tassan, J. P., Nigg, E. A. & Schibler, U. (1995). *Cell* **83,** 137-46.

21. Yankulov, K., Todorov, I., Romanowski, P., Licatalosi, D., Cilli, K., McCracken, S., Laskey, R. & Bentley, D. L. (1999). *Mol. Cell. Biol.* **19,** 6154-63.

22. Chao, D. M., Gadbois, E. L., Murray, P. J., Anderson, S. F., Sonu, M. S., Parvin, J. D. & Young, R. A. (1996). *Nature* **380,** 82-5.

23. Wang, Z., Luo, T. & Roeder, R. G. (1997). *Genes Dev.* **11,** 2371-82.

24. Grummt, I., Roth, E. & Paule, M. R. (1982). *Nature* **296,** 173-4.

25. Brow, D. A. & Guthrie, C. (1990). *Genes Dev.* **4,** 1345-56.

26. Wang, C. K. & Weil, P. A. (1989). *J. Biol. Chem.* **264,** 1092-9.

27. Schultz, P., Marzouki, N., Marck, C., Ruet, A., Oudet, P. & Sentenac, A. (1989). *EMBO J* **8,** 3815-24.

28. Joazeiro, C. A., Kassavetis, G. A. & Geiduschek, E. P. (1996). *Genes Dev.* **10,** 725-39.

29. Kassavetis, G. A., Joazeiro, C. A., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S. & Blanco, J. A. (1992). *Cell* **71,** 1055-64.

30. Gerlach, V. L., Whitehall, S. K., Geiduschek, E. P. & Brow, D. A. (1995). *Mol. Cell. Biol.* **15,** 1455-66.

31. Roberts, S., Colbert, T. & Hahn, S. (1995). *Genes Dev.* **9,** 832-42.

32. Chedin, S., Ferri, M. L., Peyroche, G., Andrau, J. C., Jourdain, S., Lefebvre, O., Werner, M., Carles, C. & Sentenac, A. (1998). *Cold Spring Harb. Symp. Quant. Biol.* **63,** 381-9.

33. Dumay-Odelot, H., Acker, J., Arrebola, R., Sentenac, A. & Marck, C. (2002). *Mol. Cell. Biol.* **22,** 298-308.

34. Huet, J., Conesa, C., Manaud, N., Chaussivert, N. & Sentenac, A. (1994). *Nucleic Acids Res.* **22,** 3433-9.

35. Eisenmann, D. M., Dollard, C. & Winston, F. (1989). *Cell* **58,** 1183-91.

36. Colbert, T. & Hahn, S. (1992). *Genes Dev.* **6,** 1940-9.

37. Kassavetis, G. A., Nguyen, S. T., Kobayashi, R., Kumar, A., Geiduschek, E. P. & Pisano, M. (1995). *Proc. Natl. Acad. Sci. USA* **92,** 9786-90.

38. Kassavetis, G. A., Braun, B. R., Nguyen, L. H. & Geiduschek, E. P. (1990). *Cell* **60,** 235-45.

39. Ferri, M. L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C. & Sentenac, A. (2000). *Mol. Cell. Biol.* **20,** 488-95.

40. Colbert, T., Lee, S., Schimmack, G. & Hahn, S. (1998). *Mol. Cell. Biol.* **18,** 1682-91.

41. Kumar, A., Kassavetis, G. A., Geiduschek, E. P., Hambalko, M. & Brent, C. J. (1997). *Mol. Cell. Biol.* **17,** 1868-80.

42. Kassavetis, G. A., Kumar, A., Ramirez, E. & Geiduschek, E. P. (1998). *Mol. Cell. Biol.* **18,** 5587-99.

43. Kassavetis, G. A., Kumar, A., Letts, G. A. & Geiduschek, E. P. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 9196-201.

44. Shah, S. M., Kumar, A., Geiduschek, E. P. & Kassavetis, G. A. (1999). *J. Biol. Chem.* **274,** 28736-44.

45. Persinger, J., Sengupta, S. M. & Bartholomew, B. (1999). *Mol. Cell. Biol.* **19,** 5218-34.

46. Kassavetis, G. A., Bardeleben, C., Kumar, A., Ramirez, E. & Geiduschek, E. P. (1997). *Mol. Cell. Biol.* **17,** 5299-306.

47. Buratowski, S., Hahn, S., Sharp, P. A. & Guarente, L. (1988). *Nature* **334,** 37-42.

48. Burley, S. K. & Roeder, R. G. (1996). *Annu. Rev. Biochem.* **65,** 769-99.

49. Margottin, F., Dujardin, G., Gerard, M., Egly, J. M., Huet, J. & Sentenac, A. (1991). *Science* **251,** 424-6.

50. Cormack, B. P. & Struhl, K. (1992). *Cell* **69,** 685-96.

51. Schultz, M. C., Reeder, R. H. & Hahn, S. (1992). *Cell* **69,** 697-702.

52. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. (1989). *Cell* **58,** 1173-81.

53. Chalker, D. L. & Sandmeyer, S. B. (1992). *Genes Dev.* **6,** 117-28.

54. Schmidt, M. C., Kao, C. C., Pei, R. & Berk, A. J. (1989). *Proc. Natl. Acad. Sci. USA* **86,** 7785-9.

55. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature* **365,** 512-20.

56. Kuddus, R. & Schmidt, M. C. (1993). *Nucleic Acids Res.* **21,** 1789-96.

57. Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. (1993). *Proc. Natl. Acad. Sci. USA* **90,** 8174-8.

58. Coleman, R. A., Taggart, A. K., Benjamin, L. R. & Pugh, B. F. (1995). *J. Biol. Chem.* **270,** 13842-9.

59. Perez-Howard, G. M., Weil, P. A. & Beechem, J. M. (1995). *Biochemistry* **34,** 8005-17.

60. Jackson-Fisher, A. J., Burma, S., Portnoy, M., Schneeweis, L. A., Coleman, R. A., Mitra, M., Chitikila, C. & Pugh, B. F. (1999). *Biochemistry* **38,** 11340-8.

61. Lee, M. & Struhl, K. (2001). *Genetics* **158,** 87-93.

62. Daugherty, M. A., Brenowitz, M. & Fried, M. G. (1999). *J. Mol. Biol.* **285,** 1389-99.

63. Daugherty, M. A., Brenowitz, M. & Fried, M. G. (2000). *Biochemistry* **39,** 4869-80.

64. Kou, H. & Pugh, B. F. (2004). *J. Biol. Chem.* **279,** 20966-73.

65. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature* **365,** 520-7.

66. Grove, A., Galeone, A., Yu, E., Mayol, L. & Geiduschek, E. P. (1998). *J. Mol. Biol.* **282,** 731-9.

67. Liu, Y. & Schepartz, A. (2001). *Biochemistry* **40,** 6257-66.

68. Whitehall, S. K., Kassavetis, G. A. & Geiduschek, E. P. (1995). *Genes Dev.* **9,** 2974-85.

69. Cox, J. M., Hayward, M. M., Sanchez, J. F., Gegnas, L. D., van der Zee, S., Dennis, J. H., Sigler, P. B. & Schepartz, A. (1997). *Proc. Natl. Acad. Sci. USA* **94,** 13475-80.

70. Hoopes, B. C., LeBlanc, J. F. & Hawley, D. K. (1992). *J. Biol. Chem.* **267,** 11539-47.

71. Petri, V., Hsieh, M., Jamison, E. & Brenowitz, M. (1998). *Biochemistry* **37,** 15842-9.

72. Campbell, K. M., Ranallo, R. T., Stargell, L. A. & Lumb, K. J. (2000). *Biochemistry* **39,** 2633-8.

73. Alexander, D. E., Kaczorowski, D. J., Jackson-Fisher, A. J., Lowery, D. M., Zanton, S. J. & Pugh, B. F. (2004). *J. Biol. Chem.* **279,** 32401-6.

74. Chaussivert, N., Conesa, C., Shaaban, S. & Sentenac, A. (1995). *J. Biol. Chem.* **270,** 15353-8.

75. Wu, J., Parkhurst, K. M., Powell, R. M. & Parkhurst, L. J. (2001). *J. Biol. Chem.* **276,** 14623-7.

76. Lebrun, A., Shakked, Z. & Lavery, R. (1997). *Proc. Natl. Acad. Sci. USA* **94,** 2993-8.

77. Davis, N. A., Majee, S. S. & Kahn, J. D. (1999). *J. Mol. Biol.* **291,** 249-65.

78. Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T. & Roeder, R. G. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 1060-4.

79. Parkhurst, K. M., Brenowitz, M. & Parkhurst, L. J. (1996). *Biochemistry* **35,** 7459-65.

80. Parkhurst, K. M., Richards, R. M., Brenowitz, M. & Parkhurst, L. J. (1999). *J. Mol. Biol.* **289,** 1327-41.

81. Starr, D. B. & Hawley, D. K. (1991). *Cell* **67,** 1231-40.

82. Starr, D. B., Hoopes, B. C. & Hawley, D. K. (1995). *J. Mol. Biol.* **250,** 434-46.

83. Parvin, J. D., McCormick, R. J., Sharp, P. A. & Fisher, D. E. (1995). *Nature* **373,** 724-7.

84. Jung, Y., Mikata, Y. & Lippard, S. J. (2001). *J. Biol. Chem.* **276,** 43589-96.

85. Petri, V., Hsieh, M. & Brenowitz, M. (1995). *Biochemistry* **34,** 9977-84.

86. Grove, A., Kassavetis, G. A., Johnson, T. E. & Geiduschek, E. P. (1999). *J. Mol. Biol.* **285,** 1429-40.

87. Kassavetis, G. A., Bartholomew, B., Blanco, J. A., Johnson, T. E. & Geiduschek, E. P. (1991). *Proc. Natl. Acad. Sci. USA* **88,** 7308-12.

88. Shen, Y., Kassavetis, G. A., Bryant, G. O. & Berk, A. J. (1998). *Mol. Cell. Biol.* **18,** 1692-700.

89. Cloutier, T. E., Librizzi, M. D., Mollah, A. K., Brenowitz, M. & Willis, I. M. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 9581-6.

90. Kassavetis, G. A., Riggs, D. L., Negri, R., Nguyen, L. H. & Geiduschek, E. P. (1989). *Mol. Cell. Biol.* **9,** 2551-66.

91. Juo, Z. S., Kassavetis, G. A., Wang, J., Geiduschek, E. P. & Sigler, P. B. (2003). *Nature* **422,** 534-9.

92. Bhargava, P. & Kassavetis, G. A. (1999). *J. Biol. Chem.* **274,** 26550-6.

93. Matsuzaki, H., Kassavetis, G. A. & Geiduschek, E. P. (1994). *J. Mol. Biol.* **235,** 1173-92.

94. Dieci, G. & Sentenac, A. (1996). *Cell* **84,** 245-52.

95. Flores, A., Briand, J. F., Gadal, O., Andrau, J. C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., Carles, C., Thuriaux, P., Sentenac, A. & Werner, M. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 7815-20.

96. Bardeleben, C., Kassavetis, G. A. & Geiduschek, E. P. (1994). *J. Mol. Biol.* **235,** 1193-205.

97. Kassavetis, G. A., Blanco, J. A., Johnson, T. E. & Geiduschek, E. P. (1992). *J. Mol. Biol.* **226,** 47-58.

98. Bartholomew, B., Durkovich, D., Kassavetis, G. A. & Geiduschek, E. P. (1993). *Mol. Cell. Biol.* **13,** 942-52.

99. Kassavetis, G. A., Letts, G. A. & Geiduschek, E. P. (2001). *EMBO J* **20,** 2823-34.

100. Ishiguro, A., Kassavetis, G. A. & Geiduschek, E. P. (2002). *Mol. Cell. Biol.* **22,** 3264-75.

101. Kassavetis, G. A., Grove, A. & Geiduschek, E. P. (2002). *EMBO J* **21,** 5508-15.

102. Kassavetis, G. A., Han, S., Naji, S. & Geiduschek, E. P. (2003). *J. Biol. Chem.* **278,** 17912-7.

103. Ferrari, R., Rivetti, C., Acker, J. & Dieci, G. (2004). *Proc. Natl. Acad. Sci. USA* **101,** 13442-7.

104. Bucher, P. (1990). *J. Mol. Biol.* **212,** 563-78.

105. Strubin, M. & Struhl, K. (1992). *Cell* **68,** 721-30.

CHAPTER 2

THE *SACCHAROMYCES CEREVISIAE* RNA POLYMERASE III RECRUITMENT FACTOR SUBUNITS BRF1 AND BDP1 IMPOSE A SEQUENCE PREFERENCE ON THE TATA-BINDING PROTEIN[1]

Introduction

The TATA-binding protein (TBP) is an integral part of all three nuclear RNA transcription systems (1). It is involved in initiating transcription by aiding in the recruitment and proper placement on the promoter of RNA polymerase. In the yeast RNA polymerase (pol) III system, TBP is found in the pol III-recruiting complex known as transcription factor (TF)IIIB, along with the TFIIB-related factor, Brf1, and the pol III-specific Bdp1 (formerly called B double prime). *In vivo*, TFIIIB is assembled onto the DNA via TFIIIC (2), but this requirement can be bypassed *in vitro* by inserting a TATA box into the promoter region (3), allowing TBP to bind the DNA and nucleate TFIIIB assembly. These TFIIIB complexes are indistinguishable *in vitro* from those assembled by TFIIIC (4).

TBP binds DNA in the minor groove and induces a bend on the order of ~80° (5-10). The speculated function of this bend is to bring the proteins involved in transcription closer together in order to allow more efficient interaction (9, 10). TBP binds the consensus sequence TATAa/tAa/tN, where N is any base (11-13), and an iterative *in vitro* selection performed using *Acanthamoeba* TBP on a double-stranded 84-base pair (bp) DNA containing 40 randomized positions yielded only four classes of TATA box and a preference for TATATAAG (14). Complex stability measurements determined using DNaseI footprinting and the frequency of a sequence appearing in the clones sequenced indicate that TBP is able to differentiate between

[1]Reproduced in part from *Nucleic Acids Research* article doi 10.1093/nar/gkl534.

A:T and T:A base pairs, even though the minor groove contains little information to allow this discrimination (14). The TATATAA sequence was found in 35 of 54 clones, while the TATAAAA sequence appeared only once (14). The K_{eq} for the TATATAA sequence is 1.1 nM, while the K_{eq} for TATAAAA is 3.7 nM (14). These differences may be accounted for by the relatively less flexible string of four adenines found at the downstream end of the less-frequently-selected TATAAAA sequence (7, 14).

A mutant TBP known as TBPm3 was isolated from yeast and found to bind to the sequence TGTAAA as well as to the wild type TATA box (15). The protein was isolated via genetic screening of TFIID (TBP from the pol II system) mutants that had the region between residues 190 and 205 mutagenized by replacing a portion of the DNA sequence with a degenerate oligonucleotide containing 8% random residues per base pair. These TBP mutants were assayed for the ability to bind to four altered *his3* TATA elements (TGTAAA, GATAAA, TAGAAA, and TATAGA) as determined by resistance to aminotriazole (AT), which competitively inhibits the product of the *his3* gene. The library of plasmids containing the TBP mutants was inserted into these four strains, and only the strain containing the TGTAAA sequence was able to grow in the presence of the inhibitor.

Four of the clones isolated were unable to grow on AT when the plasmid containing the mutated TBP was removed, and these plasmids were reintroduced into the parent strain, where they all allowed growth in the presence of the inhibitor. Plasmid shuffle complementation was used to test the mutants as the only copy of TBP in the cell. Three of these strains, containing a *URA3* plasmid with wtTBP and a *TRP1* plasmid with mutant 1, 2, or 3, were plated on 5-fluoroorotic acid (5-FOA), which selects against Ura^+ cells, and assayed for the ability to grow on 5 mM AT. All three strains grew, but were much slower compared to the wild type; strain 3,

which grew at higher concentrations of AT, was the least effective in performing the essential functions of wtTBP. Sequencing of the degenerate region of TBP in these mutants showed that all four strains contained the double mutation I194F and L205V, and mutant 3, with the highest *his3* gene activity, also contained the V203T substitution (15). These three residues are in close proximity in the folded protein (Figure 2.1) and are in a position to interact with the second base pair of the TATA box.

The specificity of binding of TBPm3 can be exploited to orient the protein unidirectionally on the DNA and allow investigation of TBP-DNA contacts at the downstream end of the TATA box. It has been previously suggested that addition of Brf1 and Bdp1 to the TBP-DNA complex alter its conformation or dynamics: i) While a missing nucleoside at the downstream end of the TATA box, coinciding with the site of TBP-mediated kinking (base pair -23), significantly enhances complex formation, the TFIIIB complex abrogates this preference, instead preferring missing nucleosides within an extended region downstream of the TATA box (16), ii) examination of TFIIIB interacting with a region upstream of the *SUP4* $tRNA^{Tyr}$ gene by photochemical crosslinking showed TBP in close proximity to the DNA minor groove, except for contacts to the DNA major groove at base pair -23 of the transcribed strand which were enhanced upon TFIIIB-DNA complex formation (17), and iii) the structure of a ternary complex composed of TBP, DNA and the primary TBP-binding domain of Brf1 revealed an exceptionally large number of interactions that bury 3230 $Å^2$ of TBP surface area (18). Therefore, the question we seek to answer is, whether the sequence preference of TBP for the downstream half of the TATA box differs from that of TFIIIB.

This was accomplished using TBPm3 and a 76-bp DNA (TBP-SXT, Figure 2.2) modified from the yeast *SUP4* $tRNA^{Tyr}$ gene to contain a mutant TATA box (TGTA) at -30 in a G+C-rich

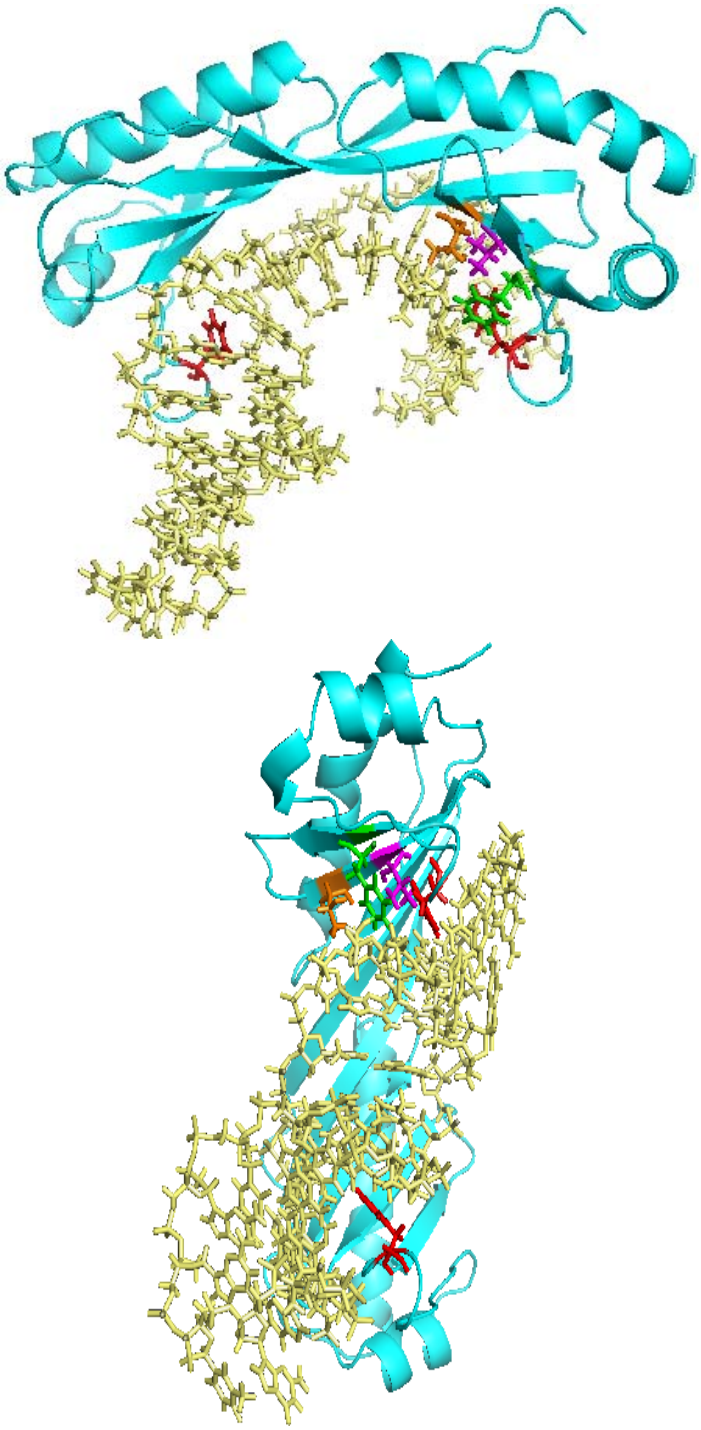

Figure 2.1. **Model of TBPm3 bound to DNA.** TBPm3 is shown in light blue and the DNA in light yellow. The amino acid changes that confer altered binding specificity are indicated (green: I194F, orange: V203T, magenta: L205V) and the intercalating phenylalanine residues are shown in red. The bottom panel shows the proximity of the mutations in the folded protein and to the second base pair of the TGTA box. This figure was generated using PyMol and a modified version of the 1YTB file from the Protein Data Bank, as described in Materials and Methods.

5'-CGC TGC AAT CTC TTT TTC AAT TGC TCC GGA CTG TAA NNN NGC GGT CCC TTA CTC TTT CCT CAA CAA TTA ACG GCC C-3'

Figure 2.2. **The TBP-SXT template used for iterative *in vitro* selection.** The first four bases of the TGTA box are underlined. N stands for any base and indicates the randomized downstream portion of the TGTA box where selection takes place.

environment, as well as four randomized bases at the downstream half of the TATA box. Here, we use an iterative in vitro selection to compare the sequence preference exhibited by TBPm3 and TFIIIB assembled with TBPm3. We show that the sequence preference of TBPm3 is less stringent than that reported for wild type (wt)TBP (14). Notably, entry of Brf1 and Bdp1 into the complex imposes a strict sequence preference for the downstream half of the TATA box that matches the TATA box of the pol III-transcribed U6 small nuclear RNA (SNR6) gene.

Materials and Methods

Generation of TBPm3-DNA model

The 1YTB file was downloaded from the Protein Data Bank and opened in InsightII molecular modeling software. One TBP-DNA complex was copied into a new file, the amino acid sequence changed to give the protein the TBPm3 sequence (I194F, V203T, and L205V), and the second base pair of the TATA box changed to a G-C pair to give the DNA the TGTA sequence. To relieve steric clashes due to the substitutions made, and obtain a more accurate model, energy minimizations were performed using the CHARMM22 force field. This minimized coordinate file was opened in PyMol, the residues colored to highlight the changes made, and the model exported to generate the figure shown.

TBPm3 purification

Protein purification was performed essentially as described previously (18), briefly: The plasmid containing the gene encoding the *Saccharomyces cerevisiae* TBPm3 (pTBPm3, a kind

gift of the Geiduschek lab at UCSD) was transformed into *E. coli* BL21(DE3)pLysS and grown to OD_{600} = 0.4 in LB broth containing 100μg/mL ampicillin. Protein overexpression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG, 1 mM, 2 hours) and the pelleted cells frozen at -80°C. Forty-five mL lysis buffer A (50 mM Tris-HCl [pH 8.0], 0.1 mM EDTA, 5% glycerol, 10 mM β-mercaptoethanol, 300 μg/mL lysozyme, and 0.5 mM phenylmethylsulfonylfluoride [PMSF]) were added to ~5 g thawed cells and incubated for 1 hour on ice. All steps after lysis are carried out at 0-4°C. The lysate was diluted ~1:1 with 60 mL lysis buffer B (50 mM Tris-HCl [pH 8.0], 1 M NaCl, 5% glycerol, 10 mM β-mercaptoethanol, and 0.5 mM PMSF), $CaCl_2$ added to 0.5 mM, and allowed to incubate for 1 hour with 10 μL DNaseI (10 U/μL). The mixture was dialyzed overnight against 3L buffer A (50 mM Tris-HCl [pH 8.0], 100 mM KCl, 20% glycerol, 1 mM EDTA, 10 mM β-mercaptoethanol, and 0.5 mM PMSF), and spun to remove precipitate prior to loading on tandem DEAE-heparin columns. TBPm3 was eluted from the heparin column using a linear gradient (100-500 mM KCl) and fractions containing TBPm3 identified by SDS-PAGE on a 12% gel. Active fractions were pooled and dialyzed against 2L buffer A for 2 hours prior to loading on a CM-Sepharose column. The protein was eluted as described above and active fractions (as determined via SDS-PAGE) were pooled and concentrated using a Centricon YM-10 (MWCO 10K). Protein concentration was determined by Coomassie Blue-staining of SDS-PAGE gels using BSA as a standard, and aliquots stored at -80°C.

Brf1 purification

The plasmid containing the gene encoding the N- and C-terminally-His-tagged *Saccharomyces cerevisiae* Brf1 (pBrf1(CNH6), also a gift of the Geiduschek lab at UCSD) was transformed into *E. coli* BL21(DE3)pLysS and grown to OD_{600} = 0.4 in LB broth containing 60μg/mL ampicillin. Protein overexpression was induced with IPTG (0.4 mM, 2 hours) and the

pelleted cells frozen at -80°C. Approximately 5 g of thawed cells were resuspended in 15 mL lysis buffer A (50 mM Tris-HCl [pH 8.0], 0.1 mM EDTA, 5% glycerol, 10 mM β-mercaptoethanol, 0.5 mM PMSF) supplemented with 1 μg/mL pepstatin and 1 μg/mL leupeptin, lysozyme was added to a final concentration of 300 μg/mL and the suspension was allowed to incubate on ice for 30 minutes. Tween-20 was added to a final concentration of 0.1%, and the cells sonicated on ice 5 times for 30 seconds. The lysate was diluted ~1:1 with 20 mL lysis buffer B (50 mM Tris-HCl [pH 8.0], 1 M NaCl, 5% glycerol, 10 mM β-mercaptoethanol, 0.5 mM PMSF) supplemented with 1 μg/mL pepstatin and 1 μg/mL leupeptin, sonicated on ice 5 times for 30 seconds, then centrifuged at 20000x*g* for 1 hour at 4°C. The pellet was resuspended in 10 mL buffer G (50 mM Tris-HCl [pH 8.0], 6 M guanidinium-HCl, 10% glycerol, 7 mM β-mercaptoethanol, 0.5 mM PMSF, 1 μg/mL pepstatin, and 1 μg/mL leupeptin), then centrifuged at 20000x*g* to pellet insoluble material. The supernatant fraction was added to 5 mL of nickel-NTA agarose beads equilibrated in buffer G and incubated for at least 1 hour on a nutator at 4°C. The beads were pelleted by centrifugation, the supernatant fraction decanted, and the beads washed 3 times for 15 minutes on the nutator at 4°C with 10 mL buffer G. After harvesting the beads by centrifugation, the protein was eluted by a pH gradient (6.7, 6.5, 5.9, 5.7, 5.5, 5.1, 4.7), accomplished via successive 15 minute washes on the nutator at 4°C with 10 mL buffer B (7 M urea, 7 mM β-mercaptoethanol, 0.5 mM PMSF, 1 μg/mL pepstatin, and 1 μg/mL leupeptin, and 100 mM sodium phosphate at the appropriate pH), and fractions containing the double His-tagged Brf1 determined via electrophoresis on a 12% SDS-PAGE gel. The active fractions were pooled, the pH adjusted to 7.9 with 1.5 M Tris-HCl [pH 8.7], and urea removed by sequential dialysis for 2 hours against 500 mL buffer C-500 (20 mM HEPES [pH 7.8], 10% glycerol, 7 mM $MgCl_2$, 500 mM NaCl, 10 mM β-mercaptoethanol, 0.01% Tween-20, 0.5 mM PMSF, 10 μM

$ZnSO_4$) containing 3.0 M, 1.5 M, 0.75 M, and 0 M urea, respectively. The sample was concentrated by centrifugation using a Centricon YM-30 (MWCO 30K), protein concentration was determined by Coomassie Blue-staining of SDS-PAGE gels using BSA as a standard, and aliquots stored at -80°C.

Rate determination for TBPm3 using electrophoretic mobility shift assays (EMSA)

An activity assay to determine the fraction of active TBPm3 recovered was performed as follows: 200 fmol of TBPm3 was incubated with 50, 200, 300, or 400 fmol of labeled dsTATA probe (see below) using the conditions described below for kinetics reactions. The reactions were subjected to electrophoresis on a 10% polyacrylamide gel, and the gel processed as described below. A plot of the quantitation data showed that all of the TBPm3 present binds DNA, as adding DNA above 200 fmol did not increase the fraction of complex formed. To determine the kinetics of TBPm3 binding to DNA, oligomers containing strong 8-bp TATA and mutant TGTA boxes were purified on denaturing polyacrylamide gels according to established protocols (19). The sequences are listed below, with the TATA and TGTA boxes bold and underlined.

$TATA_{top}$: 5'-CGT GAC TAC **TAT AAA TA**G ATG ATC CG-3'

$TATA_{bot}$: 5'-CGG ATC ATC TAT TTA TAG TAG TCA CG-3'

$TGTA_{top}$: 5'-CGT GAC TAC **TGT AAA TA**G ATG ATC CG-3'

$TGTA_{bot}$: 5'-CGG ATC ATC TAT TTA CAG TAG TCA CG-3'

For EMSA, the top strand was 5' end-labeled using T4 polynucleotide kinase and [γ^{32}P]ATP, and annealed to the bottom strand. Ten pmol of the unlabeled bottom strand was combined with 9 pmol of the unlabeled top strand and 1 pmol of the labeled top strand in a buffer containing 20 mM Tris [pH 8.0] and 50 mM NaCl. The reaction was placed in a 90°C heat block and allowed to cool slowly to room temperature. Fifty fmol of this double stranded probe was used for assays.

Reactions for kinetics assays contained 44 mM Tris [pH 8.0], 8.4 mM NaHEPES [pH 7.8], 50.5 mM NaCl, 7 mM $MgCl_2$, 1 mM EDTA, 8% (v/v) glycerol, 3 mM dithiothreitol (DTT), 4 mM β-mercaptoethanol, and 84 μg/mL bovine serum albumin (BSA). Samples were subjected to electrophoresis on native 10% polyacrylamide gels and in buffer containing 0.5 X TBE (45 mM Tris-borate [pH 8.3], 1.25 mM Na_2EDTA) and 2.5 mM $MgCl_2$.

To determine the rate of complex dissociation during electrophoresis (k_{diss}), 200 fmol TBPm3 and 50 fmol radiolabeled DNA were incubated for 55 min, 400 ng competitor DNA added [poly(dA-dT):poly(dA-dT)], and subjected to electrophoresis for time, *t*. The gels were dried, exposed to a phosphorimaging screen, and the data quantitated using ImageQuant 1.1. Data were fitted to $F_{obs} = F*\exp(-k_{diss}t)$, where F_{obs} is the observed fractional complex, F is the fractional complex present at $t = 0$, k_{diss} is the rate of dissociation on the gel, and *t* is the time of electrophoresis (20).

For determination of the off-rate in solution via EMSA, a master mix containing 750 fmol of radiolabeled DNA and 3000 fmol TBPm3 was allowed to incubate at room temperature for 1 hour, and aliquots loaded on the gel at time *t* after addition of 6000 ng poly(dA-dT):poly(dA-dT). Data were corrected by $F_{corr} = F/\exp(-k_{diss}t)$, where F_{corr} is the fractional complex corrected for dissociation during electrophoresis, F is the observed fractional complex, and *t* is the time of electrophoresis. The corrected data were fitted to $F_{corr} = F_0*\exp(-k_{off}t)$, where F_0 is the fractional complex present before addition of competitor, k_{off} is the off-rate in solution, and *t* is time after addition of poly(dA-dT):poly(dA-dT).

The on-rate in solution was determined for a protein concentration range of 20-80 nM. Protein and 50 fmol radiolabeled DNA were incubated for time *t* and immediately loaded on the gel after addition of poly(dA-dT):poly(dA-dT). The observed fractional complex was corrected

for dissociation during electrophoresis as described above and fitted to $F_{corr} = F_{final}(1-\exp(-k_{obs}t))$, where F_{final} is the calculated fractional complex present at completion of the reaction and k_{obs} the apparent first-order rate constant. The reciprocal of the slope of a plot of $1/k_{obs}$ vs. 1/[protein] yielded the second order rate constant, k_{on} (20). All rate constants represent the mean of at least three experiments.

Determination of TBPm3 and TFIIIBm3 sequence preference by iterative *in vitro* selection

The TBP-SXT oligonucleotide (5'-CGC TGC AAT CTC TTT TTC AAT TGC TCC GGA C**TG TAA** NNN NGC GGT CCC TTA CTC TTT CCT CAA CAA TTA ACG GCC C-3', mutant TATA box underlined and bold) was purified on a denaturing 5% polyacrylamide gel, and the purified single-stranded template amplified via PCR with primers PSXB (5'-GGG CCG TTA ATT GTT GAG-3') and PSXT (5'-CGC TGC AAT CTC TTT TTC-3'). The starting pool of oligonucleotides easily contains every possible sequence (4^4 or 256 sequences). The double-stranded product was 5' end-labeled using T4 polynucleotide kinase and [γ^{32}P]ATP and at least 40 ng was incubated with 405 fmol TBPm3 for 1 hour in the same buffer conditions as listed above for the kinetics studies, except with 150 mM NaCl. Incubation yielded no more than ~10% complex in the early rounds, and this fraction should yield a consensus sequence from the 256 possible sequences within 4-5 rounds of selection. After addition of 810 ng poly(dA-dT):poly(dA-dT), the reaction was loaded onto a native 10% polyacrylamide gel with the power on and subjected to electrophoresis at 175 V for 1 hour. The gel was exposed to a phosphorimaging screen, the TBPm3-DNA complex cut out of the gel, and the DNA passively eluted overnight in 1 mL elution buffer (20 mM Tris-HCl [pH 8.0], 1 M LiCl, 0.2 mM EDTA, 0.2% SDS) with rotation. The recovered DNA was amplified via PCR, the PCR product purified

on a native 7% polyacrylamide gel, radioactively labeled, and used as template for the next round of selection.

After 10 rounds of selection, the DNA was cloned into the pCR T7/NT-TOPO vector (Invitrogen) and transformed into *E. coli* TOP10. Two hundred forty clones were sequenced, 66 sequences containing the TGTAA box were aligned with ClustalX, and the frequency of each base at the 4 randomized positions scored to determine the favored sequence.

For the TFIIIBm3 selection, the selection was performed using the same conditions as for TBPm3, except that 40 ng labeled DNA was incubated with 120 fmol TBPm3, 520 fmol of total Brf1, and 1200 fmol of total Bdp1 for 1 hour, with 98.5 mM NaCl. Four hundred ng of poly(dA-dT):poly(dA-dT) was added, and the reaction was loaded onto a native 4% polyacrylamide gel with the power on and subjected to electrophoresis at 175 V for 3 hours. The recovered DNA was amplified via PCR, the PCR product purified on a native 10% polyacrylamide gel, radioactively labeled, and used as template for the next round of selection.

After 10 rounds of selection, the DNA was cloned into the pCR4-TOPO vector (Invitrogen) and transformed into *E. coli* TOP10. One hundred four clones were sequenced, 25 of which contained the TGTAA box. These were aligned and a favored sequence determined as for the TBPm3 selection.

Seventy-six- and 26-mer oligonucleotides representing a favored sequence were purified. The sequences are listed below, with the favored sequence in bold and underlined: rd10FFT 5'-CGC TGC AAT CTC TTT TTC AAT TGC TCC GGA CTG TAA **ATT G**GC GGT CCC TTA CTC TTT CCT CAA CAA TTA ACG GCC C-3', rd10FFB 5'-GGG CCG TTA ATT GTT GAG GAA AGA GTA AGG GAC CGC CAA TTT ACA GTC CGG AGC AAT TGA AAA AGA

GAT TGC AGC G-3', rd10FT 5'- CGT GAC TAC TGT AA**A TTG** ATG ATC CG-3', rd10FB 5'-CGG ATC ATC AAT TTA CAG TAG TCA CG-3'.

The 26-mers rd10FT and rd10FB were purified on 10% denaturing polyacrylamide gels (19:1), rd10FT radioactively labeled using T4 polynucleotide kinase and [γ^{32}P]ATP, and annealed to unlabeled rd10FB. Fifty fmol of the double-stranded favored sequence (dsFS) or dsTGTA were allowed to incubate with 0, 100, 200, 500, or 1000 fmol TBPm3 at room temperature for 1 hour using the same reaction conditions as described for the kinetics assays. After addition of poly(dA-dT):poly(dA-dT), the reactions were loaded onto a native 10% polyacrylamide gel with the power on and subjected to electrophoresis at 175 V for 1 hour to separate free and bound DNA. The gel was dried and exposed to a phosphorimaging screen.

Two-dimensional methidiumpropyl-EDTA (MPE)-Fe(II) footprinting of TBP-SXT sequence favored by TBPm3

The 76-mers rd10FFT and rd10FFB were purified on 5% denaturing polyacrylamide gels, rd10FFT end-labeled using T4 polynucleotide kinase and [γ^{32}P]ATP, and annealed to unlabeled rd10FFB. One hundred twenty-five fmol of the double-stranded 76-mer (dsFFS) was allowed to incubate with and without 500 fmol TBPm3 at room temperature for 1 hour using the same buffer conditions as above, except with 3 mM $MgCl_2$. After addition of 1μg poly(dA-dT):poly(dA-dT), 1 μL of 10 mM sodium ascorbate and 4 μL of 25 μM Fe-MPE were added, allowed to incubate for 1 minute, and the reaction stopped by loading onto a native 10% polyacrylamide gel with the power on and subjected to electrophoresis at 175 V for 1 hour. Free DNA and TBPm3-DNA complex were excised from the gel, and the DNA eluted and purified, as described above for the TBP-SXT selection.

The purified DNA, untreated control, and an A+G ladder were resuspended in formamide loading buffer and heated at 95°C for 2 minutes prior to loading on a 10% polyacrylamide

sequencing gel that had been pre-heated to 45°C in 1X TBE. The gel was run at 35 W in 1X TBE for ~4 hours, and dried. The gel was exposed to a phosphorimaging screen, the gel image quantitated in ImageQuant 1.1, and Microsoft Excel used to generate densitometry plots, which were overlaid in Adobe Photoshop to visualize the footprint.

Results

Purification and characterization of TBPm3

TBPm3 and Brf1 were purified using standard chromatographic techniques and concentrations determined by SDS-PAGE using BSA as a standard (Figures 2.3 and 2.4). Because of the observed differences in the ability of TBPm3 to carry out the functions of wtTBP (15), it was necessary to determine the kinetic parameters of TBPm3-DNA complex formation. This was done in order to make sure that any changes observed in sequence preference were not due to drastic differences in the DNA-binding activity of TBPm3. Since TBP, and TBPm3, binds too slowly to perform a protein titration for K_d determination (20), kinetic parameters were obtained separately using short oligonucleotides containing a strong 8-bp TATA box, or the corresponding TGTA box (Table 2.1), and used to calculate the K_d. Figures 2.5, 2.6 and 2.7 show representative gels for determination of the dissociation on the gel as a result of electrophoresis (k_{diss}), the off-rate in solution (k_{off}), and the on-rate in solution (k_{on}), respectively. Accompanying these figures are representative graphs showing fits of the data obtained from quantitation to the equations as described in Materials and Methods.

As shown in Figure 2.6, the TBPm3-DNA complex is quite stable in solution and decays with first-order kinetics, with $t_{1/2}$ of 72 and 61 minutes, respectively, for DNA containing either the TATA or TGTA box. The observed rate of dissociation is comparable to that observed for wtTBP (20, 21). Association of TBPm3 with DNA is detectable after10 s, and a gradual increase

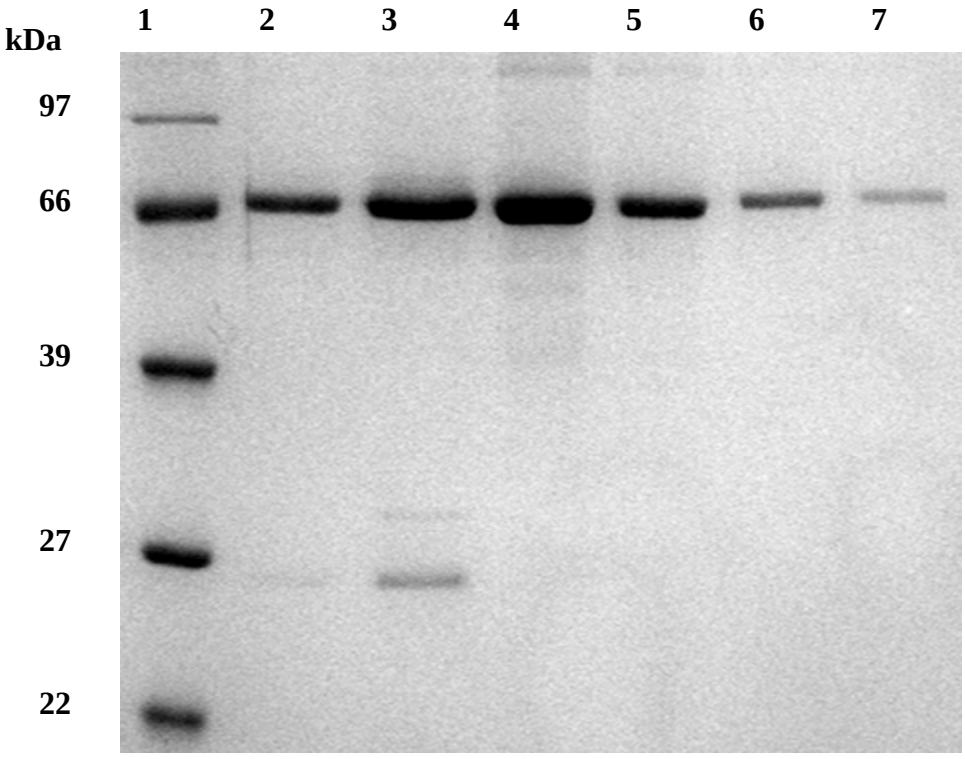

Figure 2.3. **Coomassie blue-stained 12% SDS-PAGE gel showing quantitation of purified TBPm3.** Lane 1 contains a molecular weight marker, lane 2 contains 2 μL of a 1:10 dilution of concentrated TBPm3, lane 3 contains 5 μL of the same 1:10 dilution, and lanes 4-7 contain 2.0, 1.0, 0.5, and 0.25 μg of BSA standard, respectively. TBPm3 migrates at 27 kDa, consistent with its expected molecular weight. Samples with TBPm3 contain BSA in the dilution buffer.

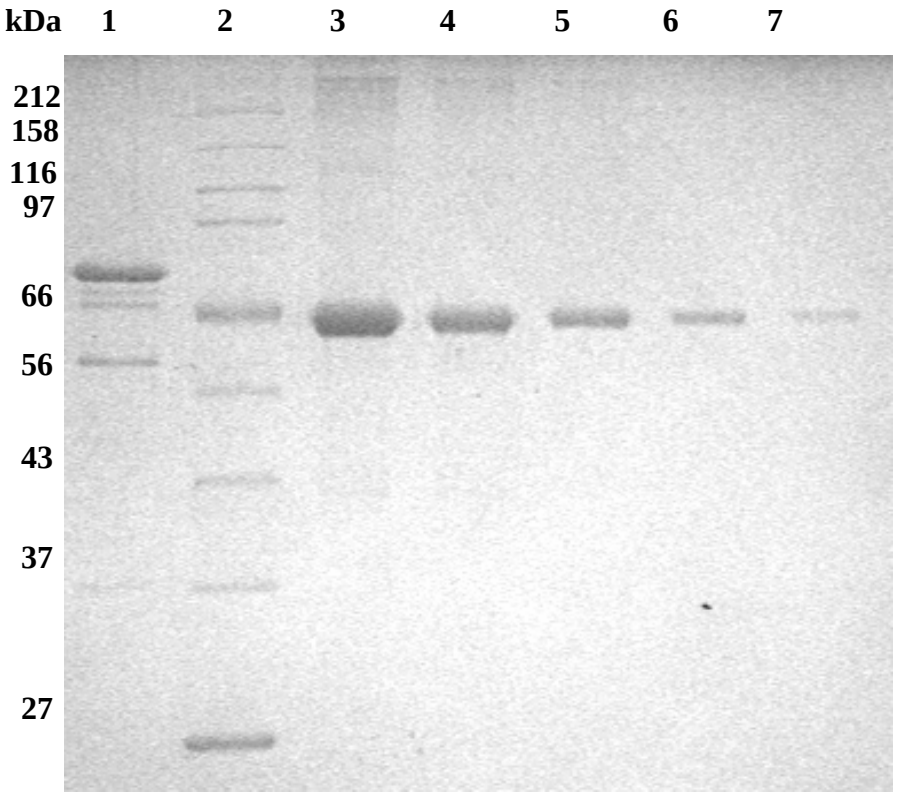

Figure 2.4. **Coomassie blue-stained 12% SDS-PAGE gel showing quantitation of purified double His-tagged Brf1.** Lane 1 contains 2 μL of concentrated Brf1, lane 2 contains a molecular weight standard, and lanes 3-7 contain 2.0, 1.0, 0.5, 0.25, and 0.125 μg of BSA standard, respectively. Brf1 migrates at ~70 kDa, consistent with its expected molecular weight of 67 kDa.

Table 2.1. **Rates of complex formation and dissociation.** Rate data determined for wtTBP and TBPm3 on two dsDNA probes with strong TATA boxes. Errors for the TBPm3 off-rates are 5% for TGTA and 11% for TATA DNA, errors for the TBPm3 on-rates are 20% for TGTA and 27% for TATA DNA, and error for the wtTBP on TATA off-rate is 9%. n.d.: not determined.

wtTBP	
TATA	**Published**
k_{diss}: 11.7 x 10^{-3} min^{-1}	k_{diss}: 5.9 x 10^{-3} min^{-1}
k_{off}: 1.4 x 10^{-4} s^{-1}	k_{off}: 1.1 x 10^{-3} s^{-1}
k_{on}: n.d.	k_{on}: 2.8 x 10^{5} M^{-1} s^{-1}
K_d: n.d.	K_d: 3.9 x 10^{-9} M
$t_{1/2}$ (solution): 85 min	$t_{1/2}$ (solution): 10 min
TBPm3	
TATA	**TGTA**
k_{diss}: 7.0 ± 2.9 x 10^{-3} min^{-1}	k_{diss}: 6.6 ± 1.3 x 10^{-3} min^{-1}
k_{off}: 1.6 x 10^{-4} s^{-1}	k_{off}: 1.9 x 10^{-4} s^{-1}
k_{on}: 5.3 x 10^{5} M^{-1} s^{-1}	k_{on}: 3.1 x 10^{5} M^{-1} s^{-1}
K_d: 3.0 x 10^{-10} M	K_d: 6.0 x 10^{-10} M
$t_{1/2}$ (solution): 72 min	$t_{1/2}$ (solution): 61 min

in complex formation is observed after longer incubation times. The second-order rate constants for association of TBPm3 with DNA containing either the TATA or TGTA box are comparable (4.3 x 10^5 $M^{-1}s^{-1}$ and 3.1 x 10^5 $M^{-1}s^{-1}$, respectively, Table 2.1 and Figure 2.7C) and well within the range of values reported for association of wtTBP with various DNA substrates, using a variety of techniques (20-27). The calculated equilibrium dissociation constant for TBPm3 binding to the TATA or TGTA probe is 0.3 nM and 0.6 nM, respectively.

Determination of TBPm3 sequence preference for the downstream half of the TGTA box

The ability of TBPm3 to bind the TGTA box unidirectionally was exploited to perform an iterative *in vitro* selection on TBP-SXT (Figure 2.2), a 76-bp oligonucleotide probe with the last four bases of the TGTA box randomized. The randomized region was selected to coincide with sites at which modulation by Brf1 and Bdp1 may be expected and includes positions 6, 7, and 8 of the TBP site and one base pair downstream (16, 17). For stringency of selection, the

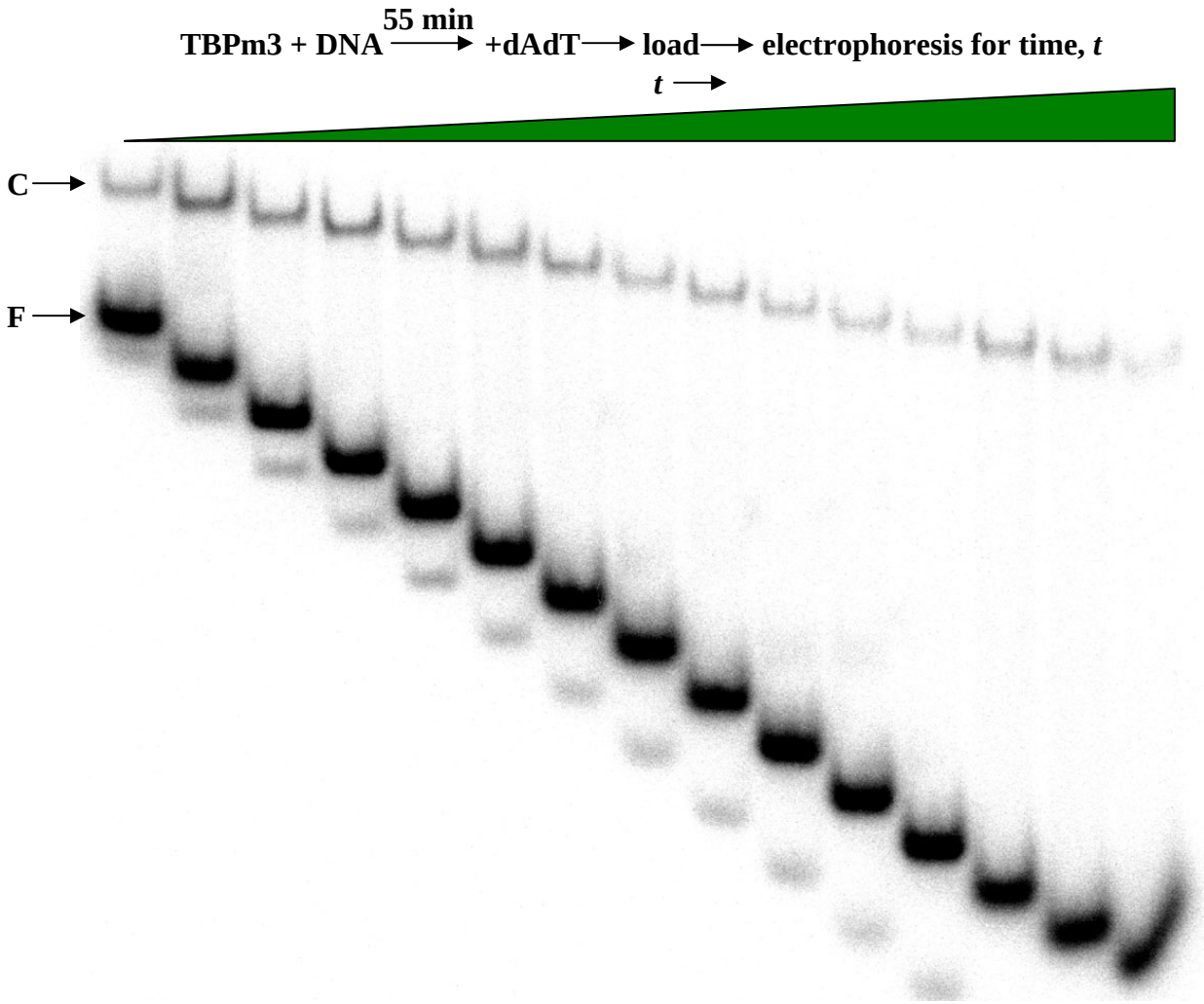

Figure 2.5A. **Determination of k_{diss} for TBPm3 on TATA DNA.** Fifty fmol of radiolabeled dsTATA 26-mer were incubated with 200 fmol TBPm3 for 55 minutes, competitor added, then electrophoresed for time, *t*. C indicates the TBPm3-DNA complex and F indicates the free dsDNA. The faint band below F is ssDNA.

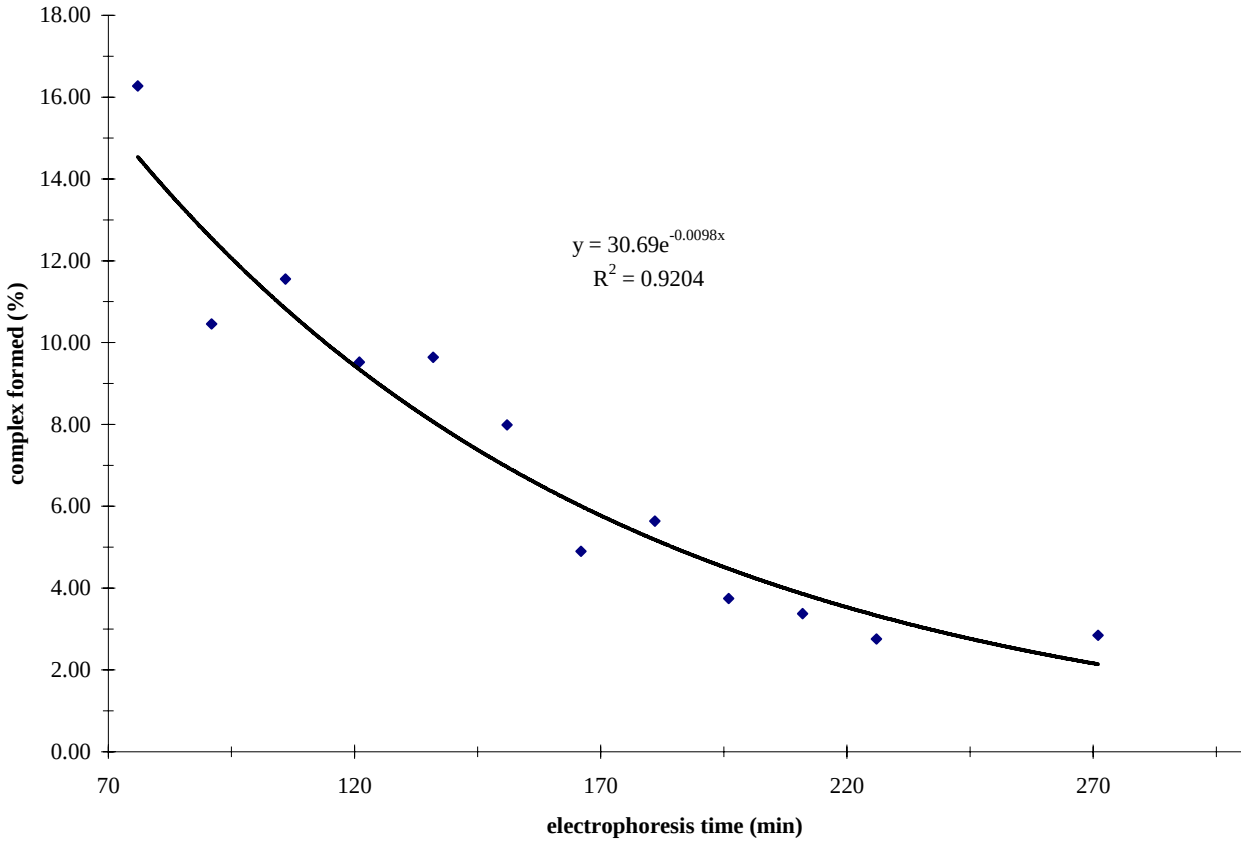

Figure 2.5B. **k_{diss} determination (TBPm3 on TATA).** After quantitation of at least three gels with ImageQuant, data were fitted to $F_{obs} = F^*\exp(-k_{diss}t)$, and the rate of dissociation on the gel extracted from the equation.

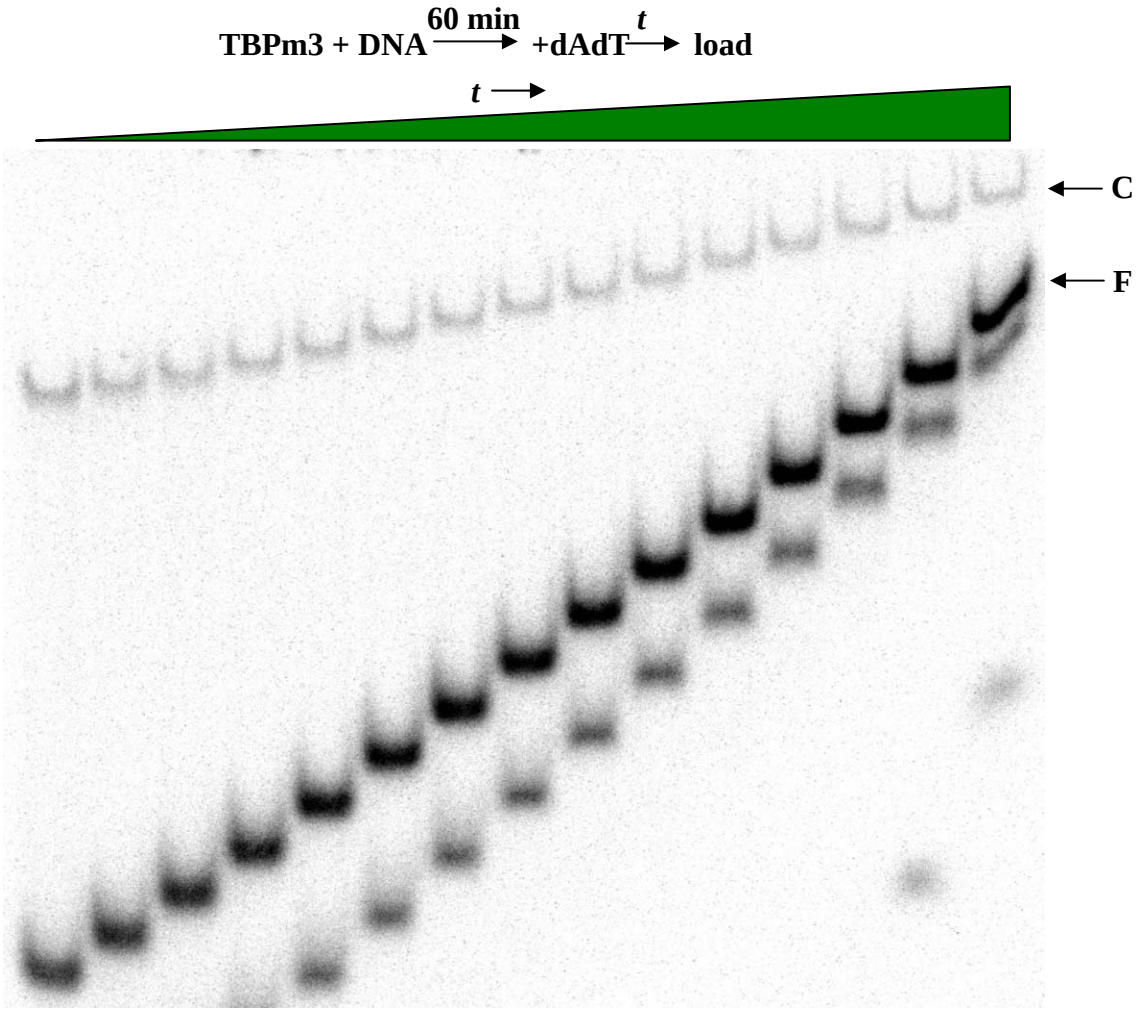

Figure 2.6A. **Determination of k_{off} for TBPm3 on TATA DNA.** A master mix containing 750 fmol of radiolabeled dsTATA 26-mer and 3000 fmol TBPm3 was allowed to incubate at room temperature for 1 hour, and aliquots loaded on the gel at time *t* after addition of poly(dA-dT):poly(dA-dT) competitor. C indicates the TBPm3-DNA complex and F indicates the free dsDNA. The faint band below F is ssDNA.

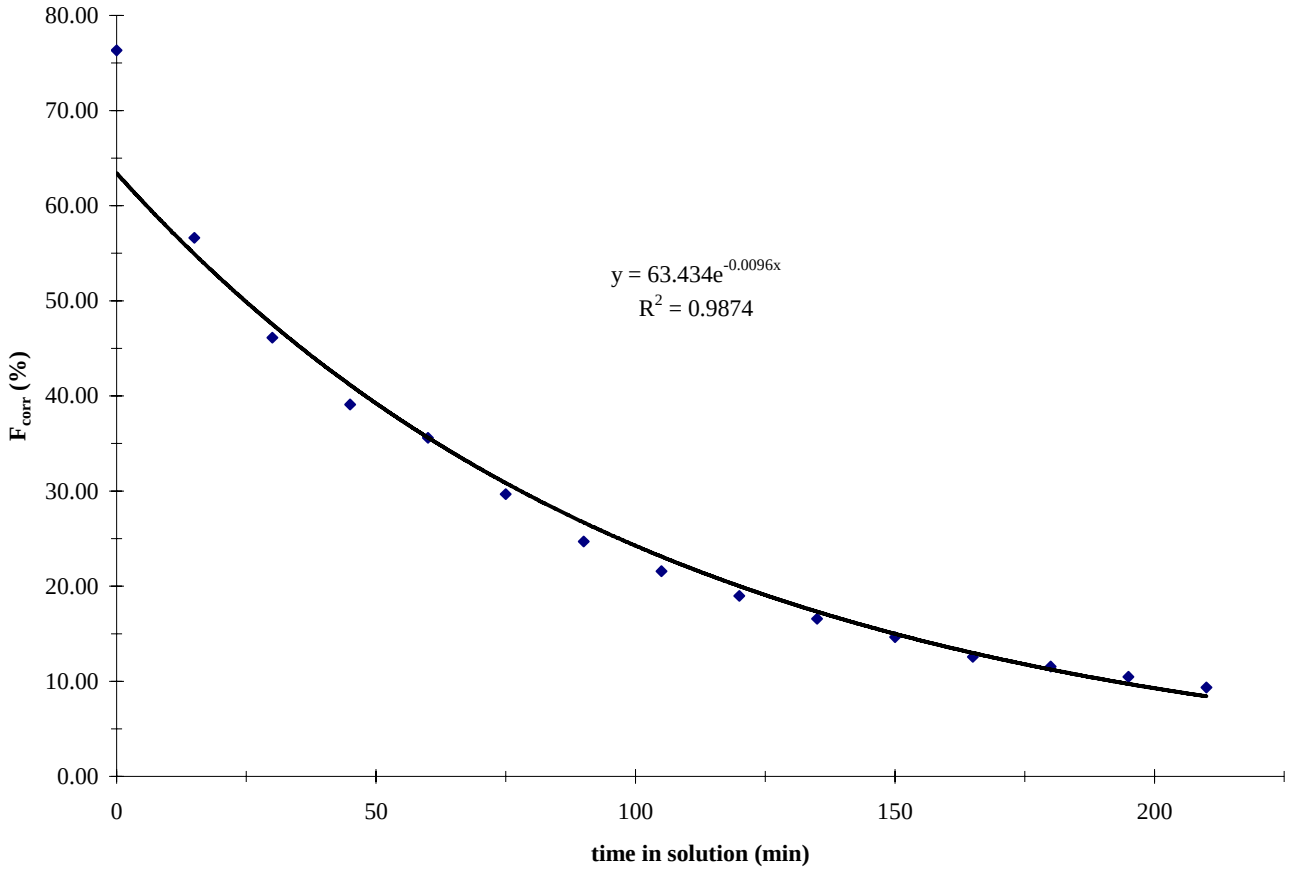

Figure 2.6B. **k_{off} determination (TBPm3 on TATA).** After quantitation of at least three gels with ImageQuant, data were corrected by $F_{corr} = F/\exp(-k_{diss}t)$, then fitted to $F_{corr} = F_0*\exp(-k_{off}t)$, and the off rate in solution extracted from the equation.

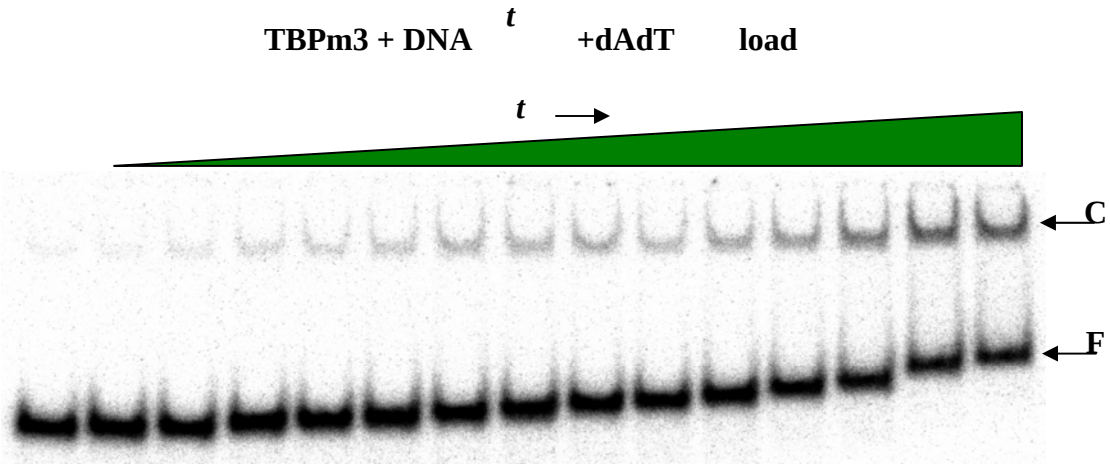

Figure 2.7A. **Determination of k_{on} for TBPm3 on TGTA DNA.** Protein (20-70 nM) and 50 fmol radiolabeled dsTGTA 26-mer were incubated for time *t* and immediately loaded on the gel after addition of competitor. C indicates the TBPm3-DNA complex and F indicates the free dsDNA.

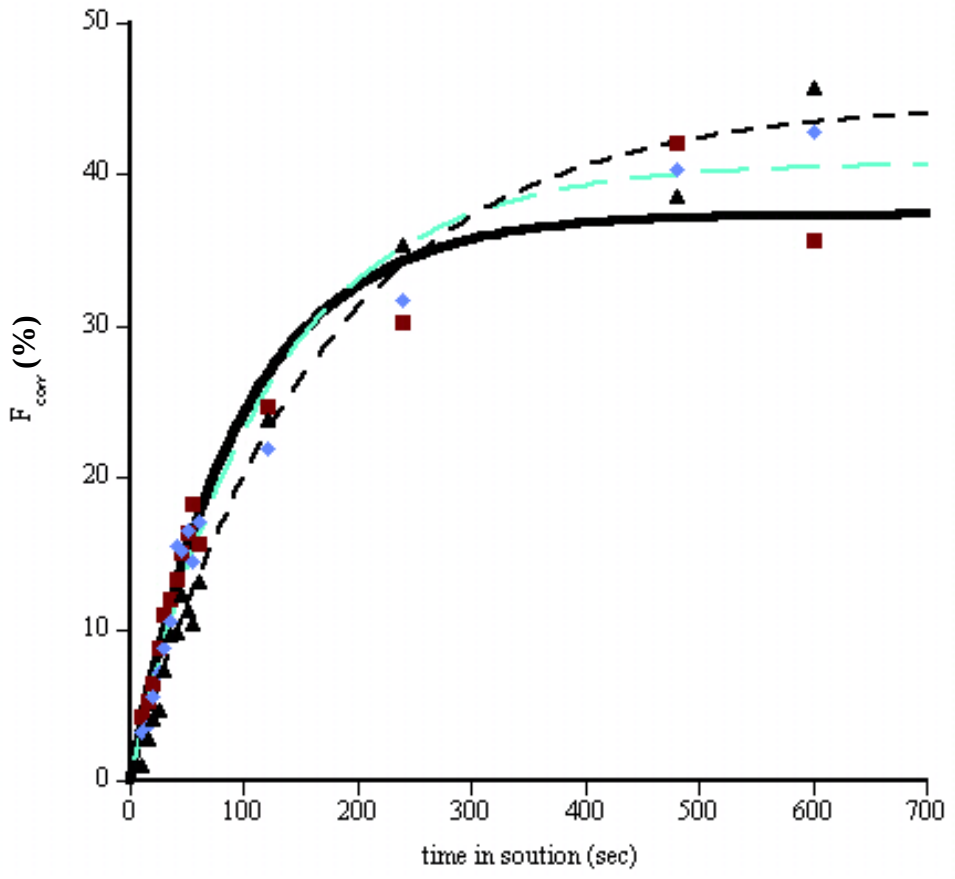

Figure 2.7B. **k_{obs} determination (TBPm3 on TGTA).** After quantitation of at least three gels (represented by three different symbols) with ImageQuant, data were corrected by $F_{corr} = F/\exp(-k_{diss}t)$, then fitted to $F_{corr} = F_{final}(1-\exp(-k_{obs}t))$.

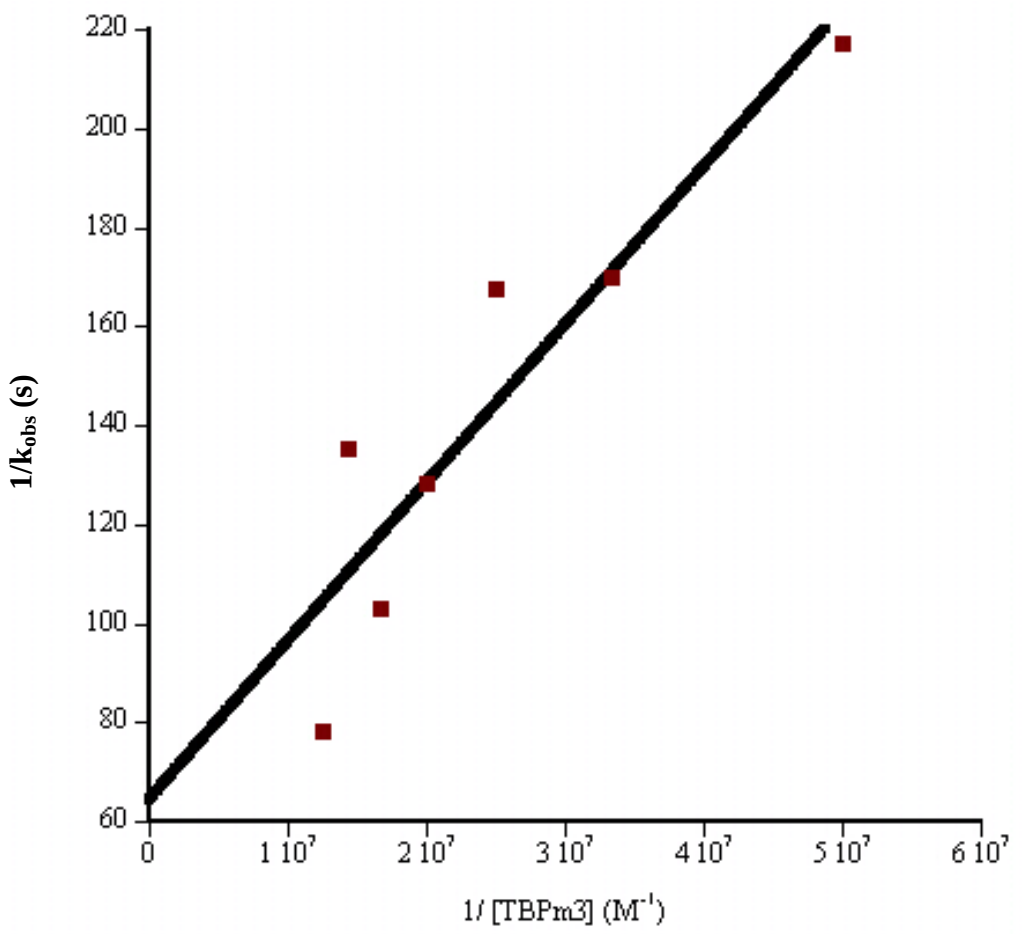

Figure 2.7C. **k_{on} determination (TBPm3 on TGTA).** The reciprocal of the slope of a plot of $1/k_{obs}$ *vs.* 1/[TBPm3] yielded the second order rate constant, k_{on}.

concentration of NaCl was raised to 150 mM during incubation of TBPm3 with DNA. With only 256 possible sequences, a consensus should be reached within 4-5 rounds of selection. However, the sequence preference exhibited by TBPm3 was only modest after five rounds of selection (data not shown), so selection was continued for another five rounds.

After 10 rounds of selection, the selected pool of DNA was sequenced and a favored sequence determined for the 4 randomized bases (Table 2.2). Two hundred forty clones were sequenced, but only 66 contained the TGTAA sequence and were used to determine the consensus for this selection. Twenty-nine clones contained a TATAA box, suggesting its

Table 2.2. **The results of sequencing after 10 rounds of selection with TBPm3.** The favored sequence of TBP-SXT (TGTAAa/gTTG) was determined by scoring the frequency of the bases at each randomized position. Over 200 clones were sequenced to obtain the 66 which contained the TGTAA box.

	N1	**N2**	**N3**	**N4**
A	23	15	15	7
C	7	6	9	8
G	20	15	17	32
T	16	30	25	19
favored	a/g	T	T	G

generation as a result of errors introduced during PCR; these sequences were not included in the alignment to avoid potential introduction of sequences arising from TBPm3 binding in the reverse orientation. Seventeen clones contained a sequence comprised of a series of GTG repeats, with only the regions complementary to the primer sequences constant. The remainder of the clones contained sequence with either no match to either of the above categories, such as other alterations to the original TGTAA sequence. Alignment of the 66 TGTAA-containing sequences still showed a surprisingly modest sequence preference for each of the randomized positions. While a C is generally disfavored at every position, position N1, corresponding to the 6^{th} base pair of the TBP site, shows essentially only selection against C. Positions 7 and 8 of the TBP site (N2 and N3) reveal a modest preference for T, while a G is preferred at position N4. This is in contrast to bases selected by wtTBP, for which a G at positions equivalent to N1 and N2 was not observed. Apparently, TBPm3 exhibits a less stringent sequence preference compared to wtTBP.

The results of the TBPm3 selection were verified by EMSA and MPE-Fe(II) footprinting on a DNA probe representing the most frequently occurring bases at each position, TGTAAATTG (note that this sequence represents the most frequently selected bases at each position, but was not found among the selected clones). TBPm3 was seen to bind to dsFS, the

26-bp DNA probe containing the selected sequence, while disfavored sequences, containing for example a C in position N1 (TGTAACTGG) yield barely detectable complex formation (Figure 2.8 and data not shown). MPE-Fe(II) footprinting on 76 bp DNA containing the favored sequence indicates that TBPm3 is binding at the TGTA box, despite the fact that this sequence was not found among the selected clones, as seen by the partial protection from cleavage at positions -28 to -23 (Figure 2.9), where the first T in the TGTA box is designated -30. Enhanced cleavage was observed at base pair -19, -18 and -31 consistent with that observed for wtTBP at these positions (16, 21).

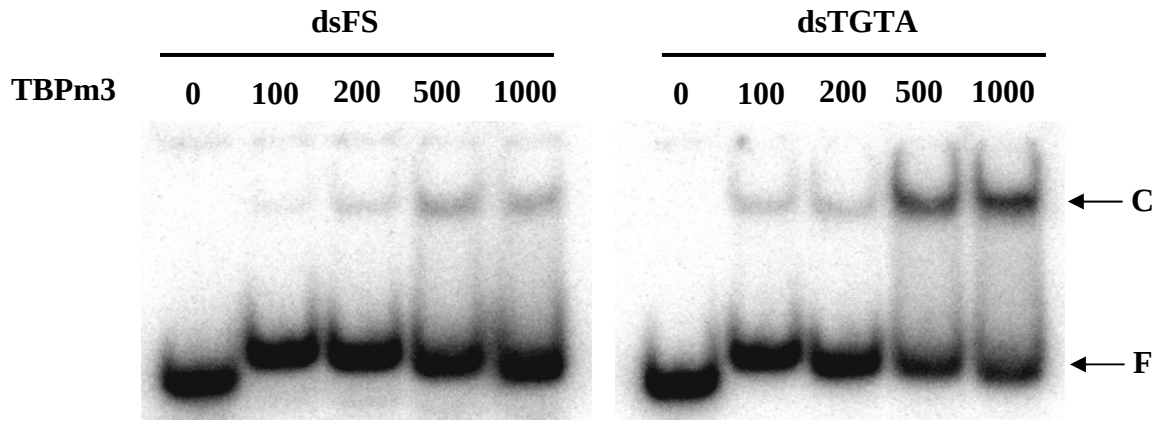

Figure 2.8. **TBPm3 binds to DNA representing the selected sequence.** Fifty fmol of each 26-mer was incubated with 0, 100, 200, 500, and 1000 fmol TBPm3. The DNA used for the left hand gel is the dsFS probe and for the right hand gel is the dsTGTA probe used in the kinetics assays. Similar levels of complex formation can be seen in both gels. C indicates the TBPm3-DNA complex and F indicates the free dsDNA.

Because introduction of Brf1 and Bdp1 into the TFIIIB complex is known to change TBP-DNA contacts at the downstream half of the TATA box (16, 17), the selection on TBP-SXT was repeated using TFIIIBm3, the TFIIIB complex assembled with a limiting amount of TBPm3. After 10 rounds of selection, 104 clones were sequenced, but only 25 contained the TGTAA

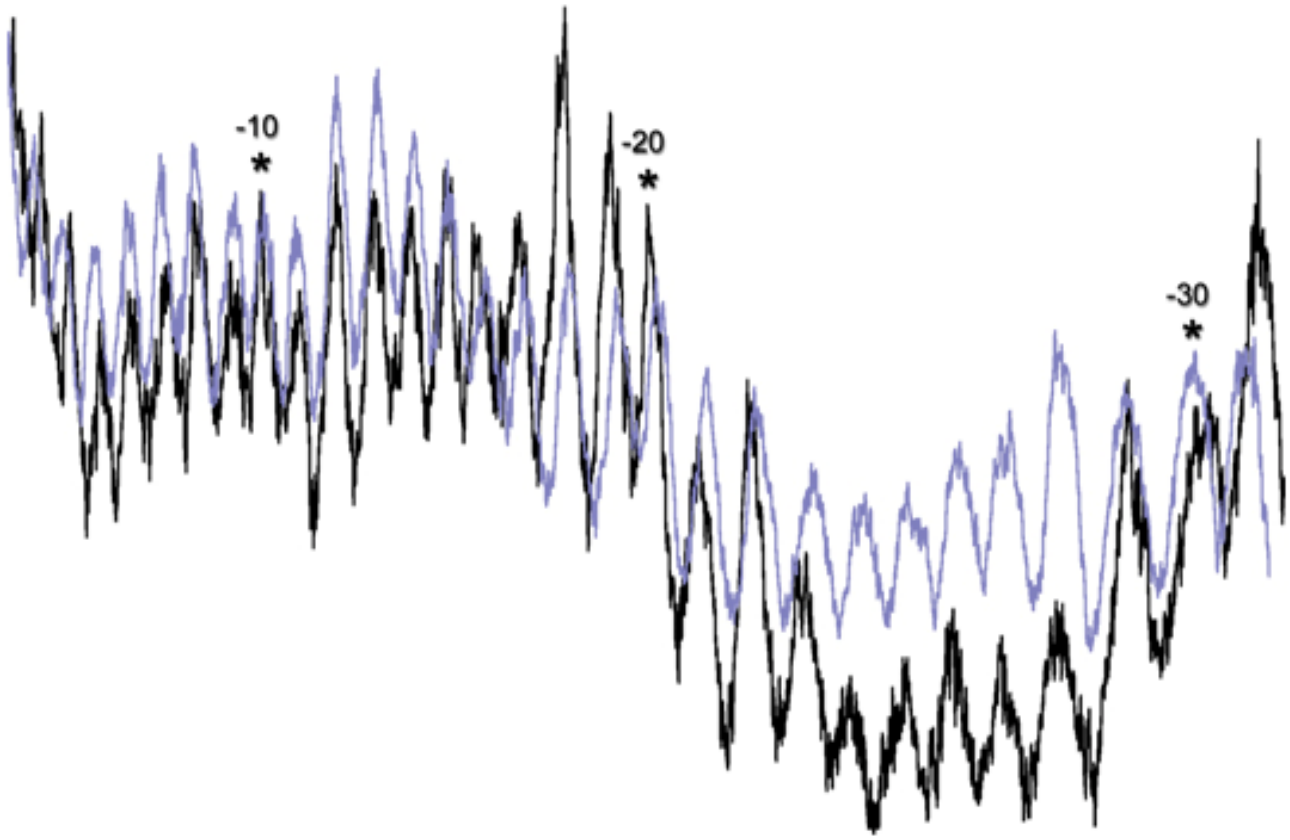

Figure 2.9. **MPE-Fe(II) 2D footprinting confirms binding of TBPm3 at the TGTA box.** Densitometry profiles of TBP-SXT containing the favored sequence (TGTAAATTG) incubated with (black line) and without (blue line) TBPm3 show protection at the TGTA box. Numbering is based on the start site of transcription (+1).

sequence and were used to determine the consensus for this selection. As for the TBPm3 selection, clones containing the TATAA sequence (25 clones) were excluded from the alignment, as were 13 clones containing the series of GTG repeats. The reasons for this radical transformation of the region between the primers are not clear, nor is it clear why TBPm3 would preferentially bind such a G+C-rich region of DNA. While the occurrence of the GTG-repeat sequences in both selections is curious, this observation was not pursued further. The remainder of the clones contained sequence with no match to the above categories, including alterations to the 5' half of the TGTA box. After 10 rounds of selection, the sequencing results (Table 2.3) showed a much stronger preference compared to TBPm3 alone, despite reaction conditions that

should have allowed less stringent binding (100 mM NaCl *vs.* 150 mM for TBPm3 selection). Notably, the selected consensus sequence (TGTAAATAG) is a perfect match to the 8 bp U6 TATA box (TATAAATA).

Table 2.3. **The results of sequencing after 10 rounds of selection with TFIIIB assembled with TBPm3.** The favored sequence of TBP-SXT (TGTAAATAG) was determined by scoring the frequency of the bases at each randomized position. Over 100 clones were sequenced to obtain the 25 which contained the TGTAA box.

	N1	**N2**	**N3**	**N4**
A	23	1	21	1
C	0	1	0	2
G	1	2	2	19
T	1	21	2	3
favored	A	T	A	G

Discussion

TBPm3-DNA complex formation

TBPm3 dissociates from its DNA site with first-order kinetics and exhibits second-order kinetics of association, as reported for wtTBP (20-27). As also seen for wtTBP under comparable experimental conditions, rate determinations do not indicate and contribution from a competing TBPm3 monomer-dimer equilibrium (21, 28). Rates of association with either DNA probe are within the range reported for wtTBP, while the rate of dissociation is slower ($t_{1/2}$ of 61 and 72 min) compared to ~10 min for wtTBP using DNA containing the 8 bp U6 TATA box (20). This difference may be due to the lower [NaCl] used here, as shown by the enhanced rate of dissociation of wtTBP that accompanies an increase in [KCl] from 60 to 80 mM ($t_{1/2}$ 100 *vs.* 65 min using the AdML promoter TATA sequence; (20)). Additionally, more stable complex formation may be the consequence of sequence flanking the 8 bp TATA box (the A immediately downstream of the U6 TATA box used in previous assays (21) was replaced with a G in the constructs used here); while TBPm3 dissociates from TATAAATAG with $t_{1/2}$ = 72 min (Table

2.1), $t_{1/2}$ for dissociation from TATAAATAA is 53 min (data not shown). Sequence flanking the TATA box has previously been shown to modulate kinetics of TBP dissociation, in particular for TBP sites characterized by alternating A-T base pairs (27, 29).

TBPm3 binds to the TGTA and TATA probes with comparable affinity, but the modestly higher affinity for the TATA-containing DNA (K_d = 0.3 nM *vs.* 0.6 nM for TGTA-containing DNA) is consistent with the identification of numerous TATA-containing sequences in the *in vitro* selections. The basis for this difference in affinity may be the increased flexibility of the T•A step relative to the T•G step (30, 31). As for wtTBP, the rate of association of TBPm3 with DNA is orders of magnitude slower than the diffusion limit; for wtTBP, the rate of association appears not to be affected by flexure at the sites of DNA kinking, whereas complex stability is (21). Consistent with this observation, rates of association of TBPm3 with either TATA- or TGTA- containing DNA are equivalent.

Sequence preference of TBPm3

The orientation of TBP on the TATA box is such that the C-proximal TBP domain interacts with the 5'-half of the TATA box, while the N-proximal domain contacts the less-conserved 3' half-site (6-8, 32). Sequence specificity at the upstream half of the TATA box has been suggested to be in part imposed by the presence of a proline (Pro191) that would disallow any base other than a T at the 5'-end of the TATA box due to steric clashes with other bases (8). The equivalent residue in the N-proximal TBP repeat is an alanine (Ala100), which imposes no such steric constraints. The modest orientational preference of TBP observed *in vitro* has been suggested to derive also from differential DNA flexure at the two sites of kinking (21). In the preinitiation complex (PIC), however, the orientation of TBP is largely determined by interaction with other transcription factors (21, 33, 34).

For TBPm3, three substitutions create a binding pocket that can accommodate a G at position 2 of the TATA box. TBPm3 exhibits an only modest sequence preference for the last four bases of the TGTA box, with C generally disfavored at every position. While A→T and T→A transversions cause little change to the chemical environment of the DNA minor groove, the introduction of GC or CG base pairs results in the exocyclic amino group of G protruding into the minor groove. For wtTBP, cavities in the interface between TBP and the DNA minor groove can be seen to accommodate a G in positions 3 and 6 of the TATA box (32). The frequent occurrence of a G at position N1 (position 6 of the TBP site) was somewhat unexpected, but this portion of the helix is flattened and unwound in the wtTBP-DNA co-crystal structure, and there likewise may not be steric clashes between TBPm3 and the DNA (note the space between the DNA and TBPm3 on the left side of the top panel in Figure 2.1; (32)). The widening of the minor groove that accompanies bending into the major groove is more difficult with GC base pairs, hence the more easily deformable TA sequence is preferred by wtTBP. Upon further scrutiny of the model in Figure 2.1, several TBPm3 side chains were seen to be in closer contact with the DNA in the TATA region than their wild type counterparts. In addition to the three mutated residues (Phe194, Thr203, and Val205), three residues that showed movement were Ser163, Val213, and Thr215. Thus, it appears that TBPm3 may feature additional contacts that may support bending of more rigid sequences.

The sequence most frequently selected by wtTBP, TATATAA is followed by a G to complete the 8 base pair TBP site, with a G or a C found at the position immediately downstream (14). This sequence is selected against in the assay used here, as the first five bases of the TGTA box (TGTAA) were held constant, thus a T in position five of the TATA box could only have arisen as a result of errors during PCR. Notably, of all the selected sequences, only one featured

the sequence TGTAT, suggesting that it is not favored by TBPm3 (unlike the sequence TATAA, which occurred in 29 of the clones, despite position two of the TBP site also not corresponding to a randomized position). Selection by wtTBP for a sequence that includes an A at position five of the TATA box was followed by the sequence A-T-A, generating the U6 TATA box TATAAATA, with a C preferred at the position immediately downstream of this 8 bp TBP site (4 of 54 clones; (14)). Eight base pair alternating TA sequences were generally followed by a G or C (12 of 54 clones). Accordingly, the presence of a G following the 8 bp TBP site preferred by TBPm3 is consistent with the preferred base following 8 bp A+T-containing sequences selected by wtTBP (14, 29). Since sequence flanking the 8 bp TBP site affects complex stability but not the rate at which TBP associates with the TATA box, flanking the A+T-rich TBP site with G+C-rich sequence may create border effects that stabilize bound TBP (27, 29).

Brf1 and Bdp1 impose a strict sequence preference on TBPm3

The sequence preference of TFIIIB assembled with TBPm3 for the downstream half of the TGTA box differs significantly from that exhibited by TBPm3 alone. While TBPm3 mainly discriminates against C in positions N1-N3, entry of Brf1 and Bdp1 into the complex imposes a strict preference for the sequence A-T-A. In both selections, a G at position N4 is preferred, although only modestly so for TBPm3. Comparable to the TBPm3 selection, no sequences occur in the TFIIIB selection with the sequence TGTAT. The possibility that TFIIIB may reverse orientation is also discounted, as a C is strongly disfavored at position N2. Accordingly, the sequence selected by TFIIIB matches that of the native U6 TATA box, except that a G immediately downstream of the 8 bp TATA box is seen in preference to the naturally occurring A.

It was previously shown that stability of TBP on a 6 bp TATA box, which is suboptimal for TFIIIB assembly, is comparable to that of an 8 bp TATA box, which efficiently supports assembly of TFIIIB (21). The significant difference between TATA box sequences must therefore be structural or dynamic adaptations to interaction with Brf1 and Bdp1. In general, the DNA bending that occurs upon association with TBP brings flanking DNA segments closer together to facilitate contacts with other transcription factors that make up the PIC (18, 35-37), and sequences that promote a disposition of DNA flanking the TBP-mediated DNA bends in a direction consistent with association of Brf1 and Bdp1 may be preferred. Indeed, analysis of TBP in complex with several divergent TATA sequences reveals comparable structures, yet only some are permissible for PIC formation; base pair changes may well be tolerated in terms of binding to TBP, but may negatively affect recruitment of other transcription factors (32).

The efficiency with which the TBP-TATA complex promotes transcriptional activity depends on the sequence of the TATA box, including A-T transversions that do not alter functional groups present in the DNA minor groove. Presumably, TBP depends significantly on recognition of inherent flexibility of the TATA box, and such differences may also affect PIC assembly (32, 38). For example, molecular dynamics simulations of different TATA variants suggest that DNA flexibility is correlated with transcriptional activity by RNA pol II (39). Correlating molecular dynamics simulations of TBP-TATA complexes involving different TATA sequences with reported transcriptional activity by pol II further suggests that optimal pol II activity occurs on DNA that allows the two domains of TBP to rotate relative to each other and that allows the H2 helix of TBP to assume an optimal disposition to interact with factors that bind both TBP domains (such as Brf1). In contrast, low activity DNA sequences appear to

promote movement of the H1 helix of TBP and to involve conformational changes in the DNA (40).

TBP introduces roll deformations at either end of the TATA box (6-8). The T•A base pair step is easily deformable due to its large range of allowable roll angles and is often found in DNA sequences requiring a sharp bend (30, 31, 41). Indeed, roll deformations at the downstream kink of TATA DNA in complex with TBP vary from ~30° for A•G steps to >45° for T•A steps (41). A unique feature of the U6 TATA box sequence identified in our selections with TFIIIB is the presence of a T•A step at the downstream end of the 8 bp TBP site. While this sequence is not strongly favored by either wtTBP or TBPm3 alone, it clearly promotes formation of the TFIIIB-DNA complex. Consistent with this interpretation, *in vitro* transcription with *Drosophila* nuclear extract indicated that while pol II utilizes the TATA box TATAAAAA in the forward direction, pol III reverses orientation (42). The data suggest that the unique feature of the selected sequence is a flexibility at the downstream end of the 8 bp TATA box that promotes Brf1 and Bdp1 binding and the associated DNA deformation downstream of the TATA box (16).

References

1. Cormack, B. P. & Struhl, K. (1992). *Cell* **69,** 685-96.
2. Kassavetis, G. A., Braun, B. R., Nguyen, L. H. & Geiduschek, E. P. (1990). *Cell* **60,** 235-45.
3. Kassavetis, G. A., Joazeiro, C. A., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S. & Blanco, J. A. (1992). *Cell* **71,** 1055-64.
4. Kassavetis, G. A., Nguyen, S. T., Kobayashi, R., Kumar, A., Geiduschek, E. P. and Pisano, M. (1995). *Proc. Natl. Acad. Sci. USA* **92,** 9786-90.
5. Starr, D. B. & Hawley, D. K. (1991). *Cell* **67,** 1231-40.
6. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature* **365,** 520-7.
7. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature* **365,** 512-20.

8. Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). *J. Mol. Biol*. **261**, 239-54.

9. Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T. & Roeder, R. G. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 1060-4.

10. Starr, D. B., Hoopes, B. C. & Hawley, D. K. (1995). *J. Mol. Biol.* **250,** 434-46.

11. Bucher, P. (1990). *J. Mol. Biol.* **212,** 563-78.

12. Yamamoto, T., Horikoshi, M., Wang, J., Hasegawa, S., Weil, P. A. & Roeder, R. G. (1992). *Proc. Natl. Acad. Sci. USA* **89**, 2844-8.

13. Reddy, P. & Hahn, S. (1991). *Cell* **65**, 349-57.

14. Wong, J. M. & Bateman, E. (1994). *Nucleic Acids Res.* **22,** 1890-6.

15. Strubin, M. & Struhl, K. (1992). *Cell* **68,** 721-30.

16. Grove, A., Kassavetis, G. A., Johnson, T. E. & Geiduschek, E. P. (1999). *J. Mol. Biol.* **285,** 1429-40.

17. Persinger, J., Sengupta, S. M. & Bartholomew, B. (1999). *Mol. Cell. Biol.* **19,** 5218-34.

18. Juo, Z. S., Kassavetis, G. A., Wang, J., Geiduschek, E. P. & Sigler, P. B. (2003). *Nature* **422**, 534-9.

19. Colbert, T. & Hahn, S. (1992). *Genes Dev*. **6**, 1940-9.

20. Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

21. Hoopes, B. C., LeBlanc, J. F. & Hawley, D. K. (1992). *J. Biol. Chem.* **267**, 11539-47.

22. Grove, A., Galeone, A., Yu, E., Mayol, L. & Geiduschek, E. P. (1998). *J. Mol. Biol.* **282**, 731-9.

23. Perez-Howard, G. M., Weil, P. A. & Beechem, J. M. (1995). *Biochemistry* **34**, 8005-17.

24. Petri, V., Hsieh, M. & Brenowitz, M. (1995). *Biochemistry* **34**, 9977-84.

25. Parvin, J. D., McCormick, R. J., Sharp, P. A. & Fisher, D. E. (1995). *Nature* **373**, 724-7.

26. Parkhurst, K. M., Brenowitz, M. & Parkhurst, L. J. (1996). *Biochemistry* **35**, 7459-65.

27. Petri, V., Hsieh, M., Jamison, E. & Brenowitz, M. (1998). *Biochemistry* **37**, 15842-9.

28. Wolner, B. S. & Gralla, J. D. (2001). *J. Biol. Chem*. **276**, 6260-6.

29. Coleman, R. A. & Pugh, B. F. (1997). *Proc. Natl. Acad. Sci. USA*, **94**, 7221-6.

30. Faiger, H., Ivanchenko, M., Cohen, I. & Haran, T. E. (2006). *Nucleic Acids Res*. **34**, 104-19.

31. Calladine, C. R. & Drew, H. R. (1986). *J. Mol. Biol*. **192**, 907-18.

32. Goodsell, D. S., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1994). *J. Mol. Biol*. **239**, 79-96.

33. Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T. & Burley, S. K. (1999). *Genes Dev*. **13**, 3217-30.

34. Whitehall, S. K., Kassavetis, G. A. & Geiduschek, E. P. (1995). *Genes Dev*. **9**, 2974-85.

35. Cox, J. M., Hayward, M. M., Sanchez, J. F., Gegnas, L. D., van der Zee, S., Dennis, J. H., Sigler, P. B. & Schepartz, A. (1997). *Proc. Natl. Acad. Sci. USA* **94**, 13475-80.

36. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995). *Nature* **377**, 119-28.

37. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. (1996). *Nature* **381**, 127-51.

38. Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. (1996). *Science* **272**, 830-6.

39. Wu, J., Parkhurst, K. M., Powell, R. M., Brenowitz, M. & Parkhurst, L. J. (2001). *J. Biol. Chem*. **276**, 14614-22.

40. Qian, X., Strahs, D. & Schlick, T. (2001). *J. Mol. Biol*. **308**, 681-703.

41. Strahs, D., Barash, D., Qian, X. & Schlick, T. (2003). *Biopolymers* **69**, 216-43.

42. Dickerson, R. E. (1998). In Sarma, R. H. & Sarma, M. H. (eds.), *Structure, motion, interaction and expression of biological macromolecules.* Adenine Press, Schenectady, NY, pp. 17-36.

43. Wang, Y. & Stumph, W. E. (1995). *Proc. Natl. Acad. Sci. USA* **92**, 8606-10.

CHAPTER 3

SEQUENCE CONTEXT EFFECTS ON TATA-BINDING PROTEIN (TBP)-INDUCED BENDING IN THE *SACCHAROMYCES CEREVISIAE* RNA POLYMERASE III TRANSCRIPTION SYSTEM

Introduction

The TATA-binding protein (TBP) is required for transcription by all three nuclear RNA polymerase (pol) transcription systems (1). It is a positively-charged, 240-amino acid, 27-kDa protein (2) composed of two homologous domains—of which the C-proximal repeat recognizes the upstream half of the TATA box—and a variable N-terminus (reviewed in 3-5). TBP binds the minor groove of its consensus sequence TATAa/tAa/tN (6-8) at a position 30 base pairs upstream of the start site of transcription (9) on pol II genes and the pol III U6 gene, and induces a bend on the order of ~80° towards the major groove (10, 11). TBP is involved in initiating transcription by aiding in the recruitment and proper placement on the promoter of RNA polymerases, and to this end, it is found in several multi-subunit polymerase recruitment complexes. In the yeast pol III system, TBP is found in the RNA pol III-recruitment factor TFIIIB, which also contains Brf1 (TFIIB-related factor) and the pol III-specific Bdp1 (formerly called B double prime). TFIIIB is assembled *in vivo* via TFIIIC (12), but this requirement can be bypassed by inserting a TATA box into the promoter region (13), allowing TBP to bind the DNA and nucleate TFIIIB assembly. TFIIIB complexes assembled this way are indistinguishable from those assembled by TFIIIC (14).

In addition to the bend induced by TBP, formation of the TFIIIB complex introduces another bend between the TATA box and the start site of transcription (15). Protein-induced DNA bending facilitates transcription by bringing the proteins involved closer together in order to allow more efficient interaction (16). This is supported by work showing that, in the pol II

system, increased TBP-induced bend angles correlate with increased complex stability and levels of transcription (17). At variance with these results, crystal structures of TBP in complex with divergent TATA variants showed conservation of structure, including comparable bend angles (18). Since TBP is involved in transcription by both pol II and pol III, and since these pols share subunits, we sought to determine whether the bend angle induced by TBP may have an effect on pol III transcription. This was addressed in two ways: i) a circular permutation assay was used to determine the protein-induced bend angle on a *SUP4* $tRNA^{Tyr}$ gene that was made TFIIIC-independent by insertion of the adenovirus major late promoter (AdMLP) TATA box at -30 in a G+C-rich sequence context, and this determination was repeated for variants containing mutations in the TATA box; ii) *in vitro* transcription assays using these DNA templates were performed to correlate the magnitude of the TBP-induced bend angle to relative levels of transcription.

The data suggest that while TBP- and TFIIIB-induced bend angles are essentially unaffected by changes to the TATA box sequence, relative levels of transcription are affected. Work by others has shown that the sequence upstream of +1 affects transcription of the *SUP4* tRNA gene *in vivo*, with deletions that remove bases immediately upstream of -15 moderately defective for transcription (19). Sequence context can also affect the affinity with which TBP binds the TATA box (20-22), as well as the binding of Brf1, with Brf1 stabilizing TBP on TATA-like sequences for which it has low affinity but not TATA boxes, to which TBP binds well (20). Finally, the polarity of the TFIIIB complex can be affected by flanking DNA (23). Contrary to observations by others of a correlation between induced bend angle and relative levels of transcription (17), that correlation is not observed here. Rather, the correlation seen here may be due to properties of a particular TATA box sequence. Since the system used here is a

modified pol III gene embedded in its native sequence context, it appears that the region 5' of the start site of transcription, or the G+C-rich sequence surrounding the TATA box, has more influence over the magnitude of the TBP-induced bend angle than the sequence of the TATA box.

Materials and Methods

Generation of pNTS, pNTS mutants and bending probes

Plasmid pET5a-Bend, the pET5a plasmid modified to contain a duplicated polylinker separated by a unique NcoI site (Figure 3.1), was used to generate bending probes as described below. A *SUP4* gene modified to contain a 6-bp TATA box at -30 in a G+C-rich region was amplified from pCJ-TA30 (a kind gift of G. A. Kassavetis and E. P. Geiduschek), and NcoI sites (underlined and in bold) inserted at each end, via mutagenic PCR with primers SUP4-NcoI-A-up (5'-TTT TT**C CAT GG**C TCC GGT GTA TAA AAG CCG-3') and SUP4-NcoI-down (5'-GTG AAT GGA G**CC ATG G**AA AAC AAA AAA ATC TCC C-3'). The upstream primer also inserted an A at the 3' end of the 6 bp TATA box of pCJ-TA30, giving it the same sequence as that found in the adenovirus major late promoter (AdMLP), but in the *SUP4* sequence context. The modified *SUP4* gene, spanning -43 to 5 bases beyond the terminator, was cloned into the NcoI site of pET5a-Bend, and clones containing the desired sequence (dubbed pNTS) were used as template for mutagenic PCR and to generate the AdMLP bending probes, as described below.

Mutagenic whole-plasmid PCR was used (with the primers listed in Table 3.1) to make substitutions to the AdMLP TATA box present in pNTS. The reactions were treated with DpnI to remove supercoiled plasmid template and used to transform competent TOP10 cells (Invitrogen). The isolated mutants, dubbed pNTS-26T, pNTS-28A26T, and pNTS-29T, were confirmed by sequencing. Purified pNTS and the pNTS mutants were each incubated at 37°C overnight with

AatII, EcoRI, NdeI, XbaI, or BglII, and restriction fragments purified by electrophoresis on a native 5% polyacrylamide gel.

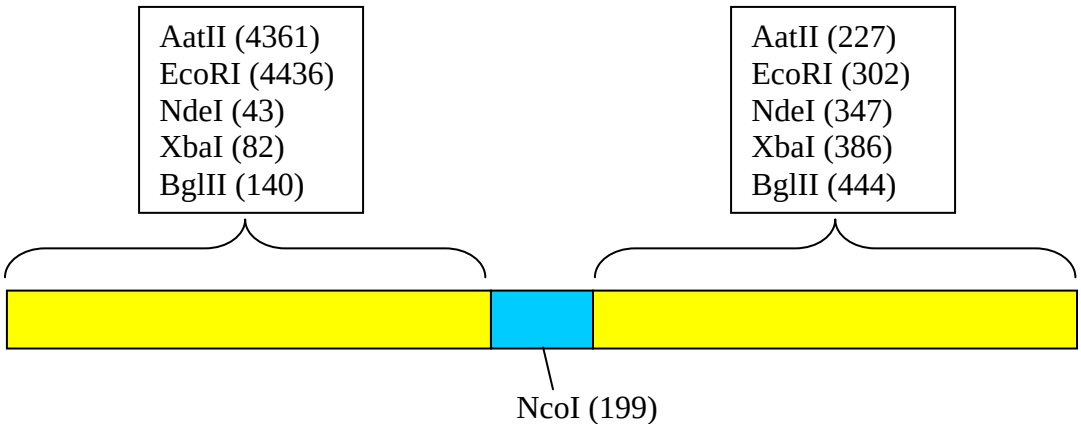

Figure 3.1. **Cartoon of a section of pET5a-Bend plasmid.** This section contains duplicated polylinker regions and the NcoI site. Digestion with enzymes that have sites in the polylinker moves the sequence inserted at the NcoI site with respect to the ends of the DNA fragment. The rest of the plasmid sequence is from pET5a.

Table 3.1. **Primers used for mutagenic whole-plasmid PCR of pNTS.** Mismatched bases used to insert changes are underlined and in bold.

pNTS	Primer name	Primer sequence
-26T	26T2	5'-ACCGCGGCTT**A**TATACACCGGA-3'
	ttd3	5'-CCCTTACTCTTTCTTCAACAATTAAATACTCTCG-3'
-29T	29T2	5'-GCGGCTTTTA**A**ACACCGGAGC-3'
	ttd4	5'-GGTCCCTTACTCTTTCTTCAACAATTAAATACTC-3'
-28A26T	28A26T2	5'-CGCGGCTT**A**T**T**TACACCGGAG-3'
	ttd5	5'-GTCCCTTACTCTTTCTTCAACAATTAAATACTCTC-3'

Determination of protein-induced bend angles in pNTS probes

The purified pNTS probes were 5' end-labeled using T4 polynucleotide kinase and [γ^{32}P]ATP. At least 20 fmol of the radioactively labeled probes were allowed to incubate with 200 fmol active TBP at room temperature for 1 hour in a reaction containing 44 mM Tris [pH 8.0], 8.4 mM NaHEPES [pH 7.8], 50.5 mM NaCl, 7 mM $MgCl_2$, 1 mM EDTA, 8% (v/v)

glycerol, 3 mM dithiothreitol (DTT), 4 mM β-mercaptoethanol, and 84 μg/ml bovine serum albumin (BSA). After addition of 400 ng poly(dA-dT):poly(dA-dT), the samples were loaded onto a native 5% polyacrylamide gel with the power on and subjected to electrophoresis at 175 V for 3.25 hours. After electrophoresis, the gels were dried, exposed to phosphorimaging screens, and gel images quantitated using ImageQuant 1.1. The procedure for determining the TFIIIB-induced bend was the same, except 156 fmol active Brf1 and 600 fmol active Bdp1 were added at the same time as TBP.

To determine the bend angle, the distance from the center of the well to the free DNA (R_{free}) and the distance from the center of the well to the TBP-DNA complex (R_{bound}) were measured in ImageQuant. The relative mobility of the DNA in complex (R_{bound}/R_{free}) was plotted *vs.* the fractional distance (D/L) of the center of the TATA box from the 5' end of the fragment, and KaleidaGraph used to fit the data to the equation 1: $R_{bound}/R_{free} = 2K(1+\cos\theta)(D/L)^2 - 2K(1+\cos\theta)(D/L) + K$ (17), where θ is the angle of the bend induced by the protein, D is the distance from the 5' end of the DNA to the vertex of the induced bend (the center of the TATA box), L is the overall length of the DNA fragment, and K is a constant. Bends induced by TFIIIB were determined by fitting the migration data to equation 2: $\mu_M/\mu_E = \cos(k\alpha/2)$, where μ_M is the mobility of the complex with the binding site at the middle of the DNA, μ_E is the mobility of the complex with the binding site at the end of the DNA, α is the induced bend angle and k is a constant.

Transcription factors

Recombinant TBP, N- and C-terminally His_6-tagged Brf1, and C-terminally His_6-tagged Bdp1 were purified and quantified as described in Chapter 2.

RNA polymerase III purification

A strain of protease deficient *Saccharomyces cerevisiae* (ATCC# 208279) was used to inoculate 9 L YPD, and the culture allowed to incubate with shaking at 30°C for 24 hours. Yeast cells were pelleted by centrifugation in a GSA rotor for 10 minutes at 4000 rpm, after which the supernatant was decanted and the cell pellets frozen at -80°C. Approximately 40 g of cells were thawed on ice (all steps following lysis are carried out on ice) in 120 mL solubilization buffer (200 mM Tris-HCl [pH 8.0], 10 mM $MgCl_2$, 10% glycerol, 10 mM β-mercaptoethanol, 0.5 mM PMSF), to which was added: ammonium sulfate to 0.5 M, β-mercaptoethanol to 10 mM, PMSF to 0.5 mM, and 2M Tris base to pH 8. The cell resuspension was lysed by three passes through a French pressure cell at ~1000 psi, pepstatin and leupeptin added to 1 μg/mL, and PMSF added to 0.5 mM. Cell debris was pelleted by ultracentrifugation for 90 minutes at 45000 rpm and 4°C. The supernatant was carefully removed and subjected to 35% ammonium sulfate precipitation. The pH was adjusted to 8.0 using 1 M NaOH, the supernatant stirred for 30 minutes on ice, and then centrifuged at 11000 rpm and 4°C for 30 minutes in a GSA rotor to pellet insoluble material. The 35% ammonium sulfate supernatant was then subjected to 70% ammonium sulfate precipitation. The pH was adjusted to 8.0 with 1 M NaOH, the supernatant stirred for 30 minutes on ice, and then centrifuged at 11000 rpm and 4°C for 30 minutes in a GSA rotor to pellet the polymerase. The supernatant was discarded and the 70% ammonium sulfate pellet resuspended in 50 mL buffer Q (50 mM Tris-HCl [pH 7.9], 50 mM ammonium sulfate, 25% glycerol, 0.5 mM DTT, and 0.5 mM Na_2EDTA). The resuspended pellet was dialyzed overnight at 4°C against 1 L buffer Q, then diluted with TGED (50 mM Tris-HCl [pH 7.9], 25% glycerol, 0.5 mM DTT, and 0.5 mM Na_2EDTA) until the conductivity matched that of buffer Q. The diluted resuspension was centrifuged in a GSA rotor for 10 minutes at 5000 rpm and 4°C to pellet any insoluble

material, pepstatin and leupeptin added to 1 μg/mL, PMSF added to 0.5 mM, and β-mercaptoethanol added to 10 mM. This was then loaded onto a DEAE-Sephadex (A25) column that had been equilibrated in buffer Q at 0.6 mL/min. The column was washed overnight with buffer Q at 0.6 mL/min, and eluted with a linear gradient of ammonium sulfate (50-425 mM) at 0.6 mL/min, with 6 mL fractions collected.

Fractions containing pol III activity were identified by using 50 μL in a 100 μL *in vitro* transcription reaction containing 50 mM Tris-HCl [pH 8.0], 1.6 mM $MnCl_2$, 1.9 μg sonicated salmon sperm DNA, 200 μM ATP, 100 μM CTP, 100 μM GTP, and 15 nM [α^{32}P] UTP. Reactions were incubated at 30°C for 30 minutes and then cooled rapidly to 0°C to stop the reaction. Five μL of each reaction was spotted on DE-81 filter paper, and unincorporated label separated from RNA by washing the filters several times with ice-cold 0.5 M Na_2HPO_4 [pH 7.0]. Filters were washed briefly with 95% ethanol, allowed to dry, and exposed to a phosphorimaging screen overnight. ImageQuant 1.1 was used to quantitate the amount of radioactive counts present on each filter. Peaks 98-105 were pooled and concentrated by centrifugation using a Centricon YM-10 (MWCO 10K), and aliquots stored at -80°C. The activity of the pooled pol III was confirmed in a 20 μL *in vitro* transcription reaction containing 50 mM Tris-HCl [pH 8.0], 1.6 mM $MnCl_2$, 382 ng sonicated salmon sperm DNA, 200 μM ATP, 100 μM CTP, 100 μM GTP, 63 nM [α^{32}P] UTP, 0.5 μg/mL α-amanitin, and 1 μL of the pooled pol III. The reaction was spotted on a DE-81 filter and processed as above. Fractional activity of the purified pol III was determined by comparison to a previously characterized preparation (a generous gift of G. A. Kassavetis and E. P. Geiduschek).

In vitro transcription with pNTS probes

Conditions for *in vitro* transcription were as follows: 44 mM Tris [pH 8.0], 8.4 mM NaHEPES [pH 7.8], 81 mM NaCl, 7 mM $MgCl_2$, 1 mM EDTA, 8% (v/v) glycerol, 3 mM dithiothreitol (DTT), 4 mM β-mercaptoethanol, 84 μg/ml bovine serum albumin (BSA), 100 fmol TBP, 500 fmol active Brf1, 600 fmol active Bdp1, at least 2.5 fmol pol III, 100 ng of pNTS (or mutant pNTS) plasmid DNA, 198 μM ATP, 99 μM CTP, 99 μM GTP, 25 μM UTP, and 42 nM [α^{32}P] UTP. Everything except the ribonucleotides was combined in a final reaction volume of 20 μL and incubated at room temperature for 30 minutes. The ribonucleotide mixture was added and transcription allowed to proceed for 30 minutes at room temperature. Stop solution (10 mM Tris-HCl [pH 8.0], 3 mM EDTA, 0.2% SDS), recovery marker, and 100 μg tRNA were added, and the reactions extracted with phenol:chloroform:IAA, followed by ethanol precipitation with ammonium acetate. Precipitated RNA was resuspended in TE' (20 μL), 2 volumes of formamide loading buffer were added (formamide with 10 mM EDTA [pH 8.0]), and the samples heated for 3 minutes at 90°C before being loaded onto a 5% denaturing polyacrylamide gel and subjected to electrophoresis for 1 hour at 600V in 0.5X TBE (45 mM Tris-borate [pH 8.3], 1.25 mM Na_2EDTA). After electrophoresis, the gel was dried, exposed to phosphorimaging screens, and gel images quantitated using ImageQuant 1.1.

Results

Determination of protein-induced bend angles

A circular permutation assay was performed to determine the magnitude of the bend induced by TBP alone and by the TFIIIB complex on DNA containing a 7 bp TATA box or a mutant TATA box. Figure 3.2A shows a representative gel used to determine TBP-induced bend angles for the four different sequences of TATA box used. As seen previously (17), the mobility

of the complex on the gel varies with the distance of the TATA box from the 5' end of the DNA, with probes where the TATA box is in the center yielding complexes that migrate more slowly than complexes formed with probes where the TATA box is at the end of the DNA (compare lanes 4 and 1). The difference in migration observed between the free DNA cut with AatII and BglII (compare lanes 1 and 5) may reflect an inherent flexibility in the DNA. Since it is observed for pNTS and all three derivatives, however, it must due to sequence outside the TATA box—perhaps the A+T-rich terminator. The difference in labeling intensity observed for some probes (compare lanes 6 and 7) does not affect the determined bend angle, as measurements of the distance from the center of the well to the center of the free DNA band at a lower intensity than shown agreed with measurements made at the intensity shown. Figure 3.2B shows a representative fit of the data obtained to the equation described in Materials and Methods. Surprisingly, fits of the data for all probes yielded angles that are very similar (Table 3.2), in contrast to those previously observed for a pol II promoter (17). In fact, the 28A26T probe, based on the sequence that was bent the least in the previous study, was shown in this work to be more bent than 26T, the probe based on the sequence that was bent the most in the previous study.

To confirm the observation that the sequence of the TATA box had no effect on the magnitude of the TBP-induced bend angle, a gel retardation assay was performed using DNA containing the four variant TATA boxes all digested with the same enzyme (XbaI), which placed the TATA box in the same position with respect to the ends of the DNA in all four probes. Figure 3.3 shows the results of this assay, with all lanes containing TBP-DNA complexes migrating to the same position on the gel. Since the migration distance is correlated with bend angle, all of the probes are bent comparably, further reinforcing the observation that TBP-induced bend angles are independent of TATA box sequence in this sequence context.

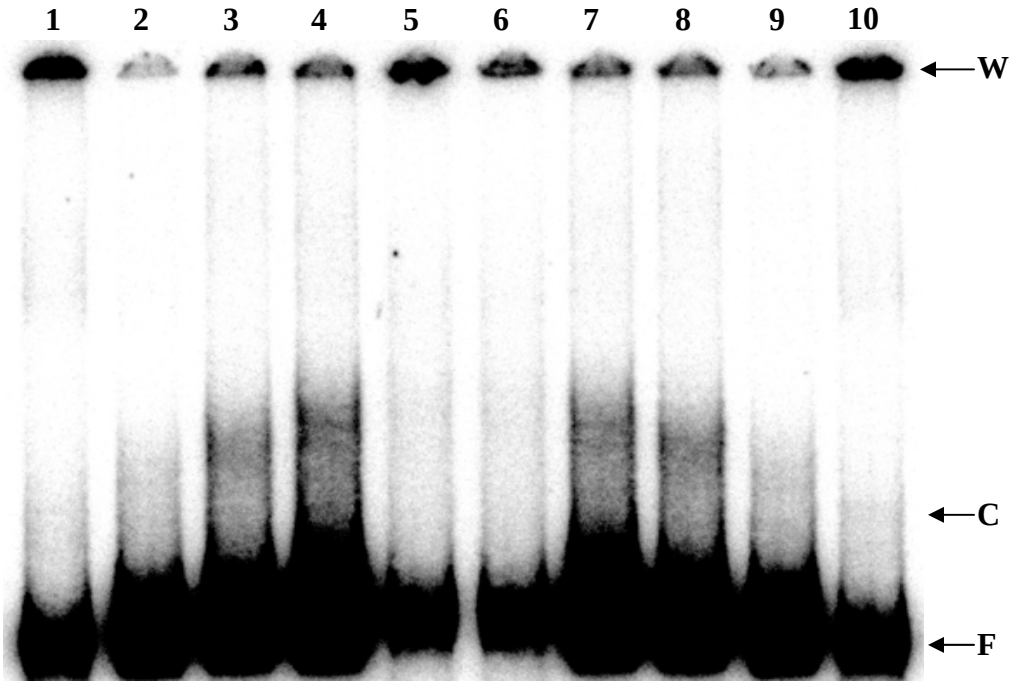

Figure 3.2A. **Determination of TBP-induced bend angle.** Lanes 1 and 10 contain pNTS-29T digested with AatII, lanes 2 and 9 with EcoRI, lanes 3 and 8 with NdeI, lanes 4 and 7 with XbaI and lanes 5 and 6 with BglII. At least 20 fmol of end-labeled probe was incubated with 200 fmol TBP and subjected to electrophoresis on a 5% polyacrylamide gel. Measurements made from the center of the well (W) to the center of the complex (C) and from the center of the well to the center of the free DNA (F) were used to calculate the relative mobility (R_{bound}/ R_{free}).

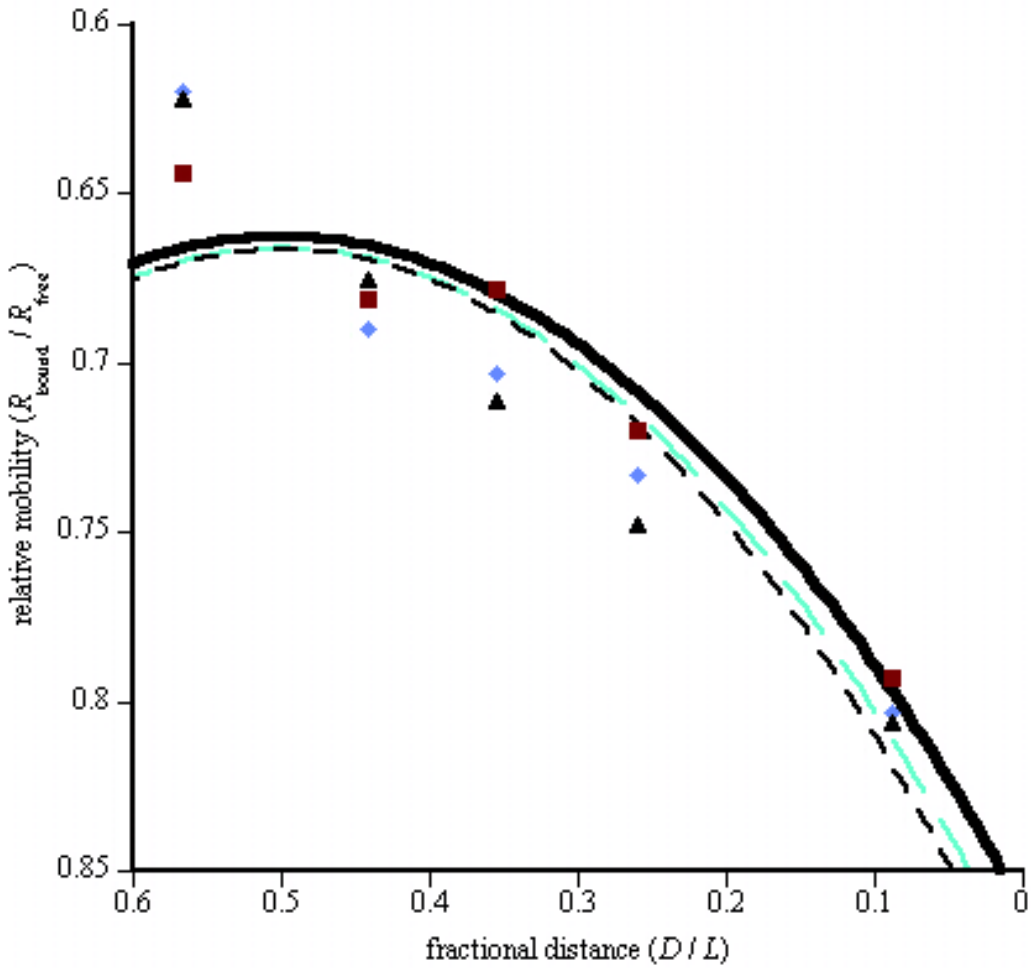

Figure 3.2B. **Relative mobility as a function of fractional distance.** The relative mobility (R_{bound}/ R_{free}) of the complex was plotted versus the fractional distance of the center of the TATA box from the 5' end of the DNA fragment (D/L). The graph shows 3 sets of data fitted to equation 1 to extract the bend angle (θ).

Table 3.2. **TBP-induced bend angles.** Sequence and corresponding bend angles of TATA boxes in probes constructed from pNTS and its derivatives. Values correspond to the average of three experiments, and errors listed are the standard deviation of the average. Differences from wild-type are underlined and in bold print. Pol II bend angles are from Starr *et al.* (17).

pNTS	Sequence	Bend angle	Pol II bend angle
-AdMLP	TATAAAA	114 ± 0°	93°
-26T	TATA**T**AA	109 ± 5°	106°
-28A26T	TA**A**A**T**AA	121 ± 3°	59°
-29T	T**T**TAAAA	121 ± 1°	87°

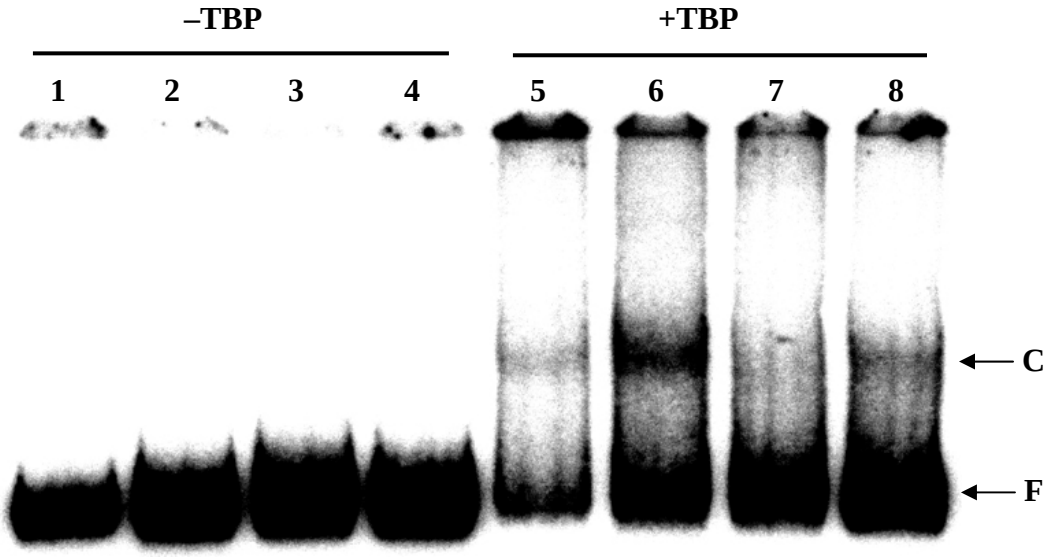

Figure 3.3. **TBP-induced bend angle is independent of DNA sequence at the TATA box.** Lanes 1 and 5 contain pNTS-AdMLP, lanes 2 and 4 contain pNTS-26T, lanes 3 and 5 contain pNTS-28A26T, and lanes 4 and 8 contain pNTS-29T, all digested with XbaI. These probes were radiolabeled, incubated with TBP and the reactions subjected to electrophoresis. The gel shows no difference in migration of the XbaI-cut probes, indicating that they are all bent to approximately the same degree. C indicates the TBP-DNA complex and F indicates the free dsDNA.

Because formation of the TFIIIB-DNA complex is known to alter the TBP-DNA contacts at the downstream half of the TATA box (15, 24), as well as introducing another bend downstream of the TATA box but before the start site of transcription (15), the bend angle for the TFIIIB-DNA complex was determined. As shown in Figure 3.4, formation of the TFIIIB-DNA complex appears to equalize the slight differences in bend angle observed with TBP alone. Fits of the 26T data to equation 1 give a bend angle of ~115°, and fits to equation 2 give a bend angle of ~120°, which was unexpected, given the additional bend induced by TFIIIB (15). These numbers agree with previous determinations of the TFIIIB-induced bend angle (25).

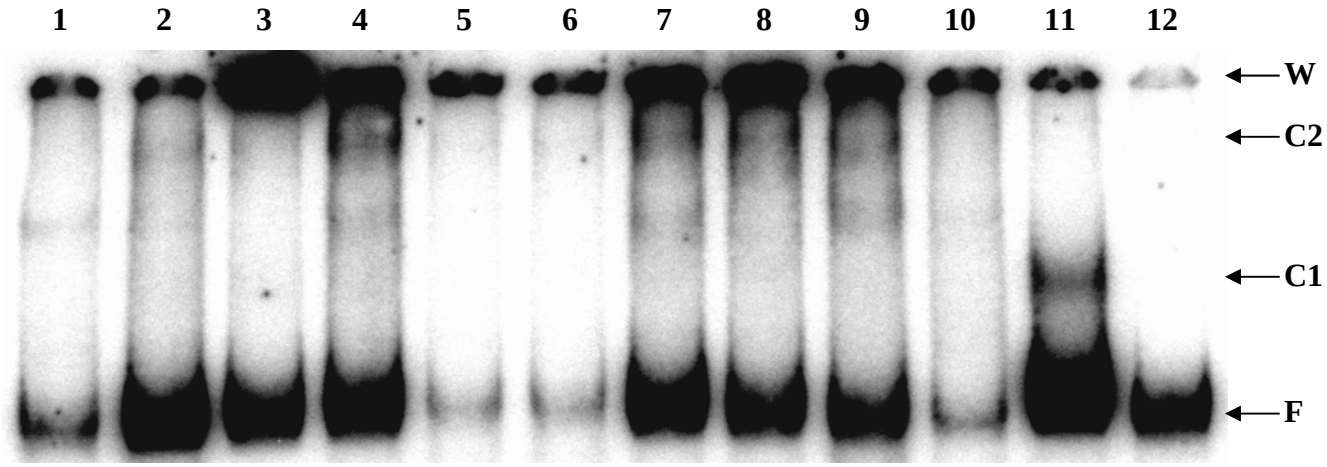

Figure 3.4. **Determination of TFIIIB-induced bend angle.** Lanes 1, 10 and 12 contain pNTS-26T digested with AatII, lanes 2 and 9 with EcoRI, lanes 3 and 8 with NdeI, lanes 4, 7 and 11 with XbaI and lanes 5 and 6 with BglII. Lane 11 contains only TBP. W marks the well, C1 indicates the TBP-DNA complex, C2 is the TFIIIB-DNA complex, and F indicates the free dsDNA.

Determination of relative transcript levels for pNTS probes

To ascertain whether TBP-induced bending had any effect on relative levels of transcription for the different sequence TATA box probes, *in vitro* transcription reactions were performed. As shown in Figure 3.5, different levels of RNA transcripts were generated with each of the variant TATA box sequences. Visual inspection of the gel revealed a difference in

transcription even between those probes that have the same bend angle (compare lanes 7-9 to lanes 10-12). Quantitation of the longest, most abundant transcript confirmed that there was no correlation between the degree of TBP-induced bending and the relative level of pol III transcription (Table 3.3). For example, although the -28A26T and -29T probes were bent to the same degree, the level of transcription relative to the -AdMLP probe, which was arbitrarily set as 1.0, varied dramatically for these two probes (1.99 *vs.* 7.62).

The bands observed in the lanes are RNA transcripts as confirmed by their degradation in reactions where RNase A was added (data not shown). These products are also dependent on the inserted TATA box, as no transcripts were detected in reactions performed with the parent pET5a-Bend plasmid (data not shown).

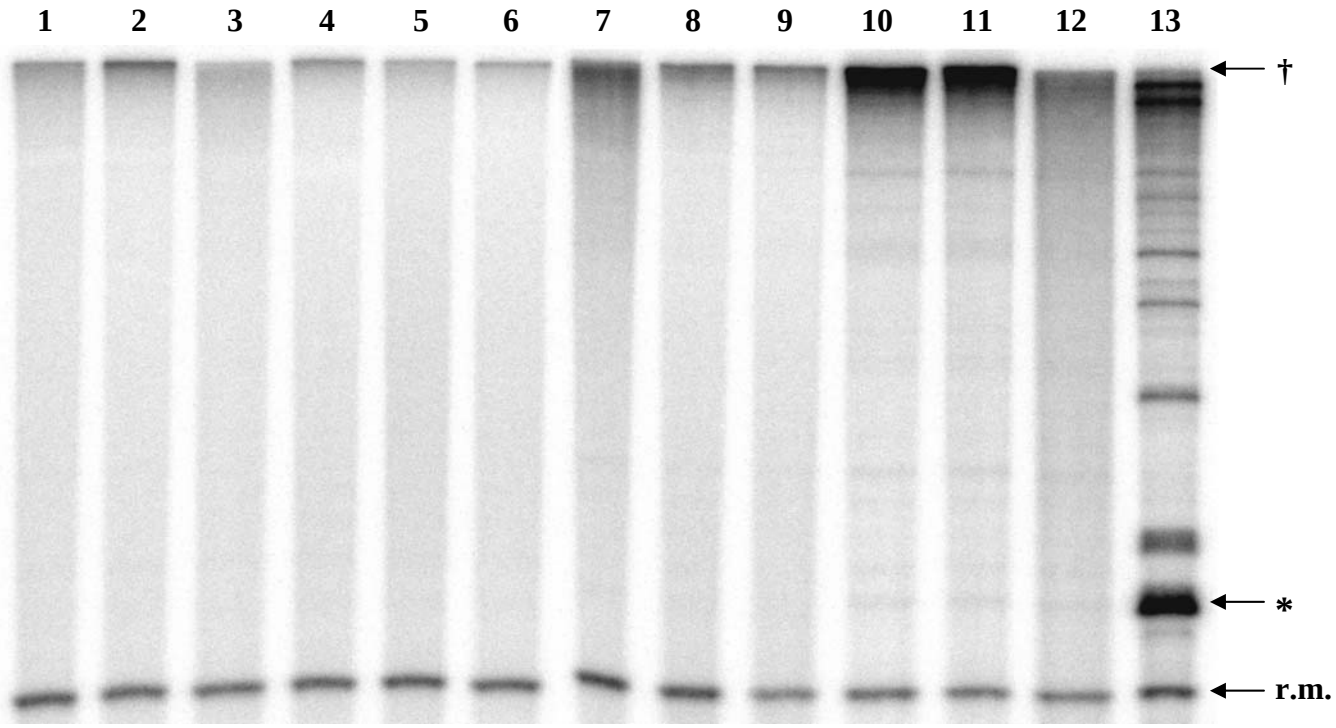

Figure 3.5. ***In vitro* transcription on pNTS and derivatives.** Lanes 1-3 contain the pNTS-AdMLP supercoiled template, lanes 4-6 contain the -26T derivative, lanes 7-9 contain the -28A26T derivative, lanes 10-12 contain the -29T derivative, and lane 13 contains the $pU6_L$boxB control. † marks the quantitated transcript, ***** marks the 100 nt transcript, and r.m. is the recovery marker.

Table 3.3. **Relative transcription levels of pNTS and derivatives.** TATA box sequence and corresponding bend angles and transcription levels of pNTS and derivatives. Values correspond to the average of three experiments, and errors listed are the standard deviation of the average. Differences from wild-type sequence are underlined and in bold print. Pol II bend angles and transcription levels are from Starr *et al.* (17).

pNTS	Sequence	Pol III Bend angle	Pol II bend angle	Pol III relative transcription	Pol II relative transcription
-AdMLP	TATAAAA	114 ± 0°	93°	1.00	1.00
-26T	TATA**T**AA	109 ± 5°	106°	0.59 ± .10	0.73
-28A26T	TA**A**A**T**AA	121 ± 3°	59°	1.99 ± .45	0.15
-29T	T**T**TAAAA	121 ± 1°	87°	7.62 ± 2.93	0.68

Notably, termination is remarkably inefficient on all pNTS derivatives; while transcription reactions with the control plasmid $pU6_L$boxB (lane 13) yields the properly terminated ~100 nt transcript in addition to the expected read-through transcripts of various lengths, little transcript is seen with the pNTS templates corresponding to accurately terminating polymerase. That the observed transcript length is a consequence of inefficient termination was confirmed by the observation that it is altered by linearizing the templates at a site upstream of the terminator (data not shown). While the pNTS templates all contain the native terminator, they include only 5 bp of *SUP4* gene sequence downstream of the terminator, followed by plasmid sequence. These data therefore suggest that sequence well beyond the required string of thymines that constitute the pol III termination signal is unexpectedly important for efficient termination.

Discussion

Protein-induced bend angles are not affected by sequence at the TATA box

Here we have used a circular permutation assay and DNA probes containing TATA boxes with varying sequence to ascertain the effect of the TATA box sequence on TBP- and

TFIIIB-induced bend angles. The methods used here are the same as those used in previous work (17) on a pol II promoter, yet almost no difference in the TBP-induced bend angle was observed with the different TATA box sequences used here. Of note is the 28A26T probe, which was bent 121° using our constructs compared to only 59° in the previous study. If the sequence at the TATA box was the main determinant of bending, then this probe should have been bent less than the pNTS parent plasmid containing the AdMLP sequence, and yet, this was not observed. It should be noted that bend angles determined using EMSA are only semiquantitative and dependent on several factors, including the choice of reaction conditions, polyacrylamide gel percentage (26), DNA sequence, and assumptions used for calculations and fitting the data to equations (17, 26). In addition, the probes used here contain an intrinsic bend, probably deriving from the A+T-rich terminator region. For all of these reasons, direct comparisons of bend angles determined for probes with the same TATA sequence by different groups may not be illustrative. However, since the bend angles in this study were all determined under the same conditions and by making changes only to the TATA box sequence, comparisons can safely be made regarding the approximate bend induced by TBP in the different probes.

The TBP-induced bend angles observed here are all essentially the same, since the resolution of the electrophoretic mobility shift assay (EMSA) technique used is not sensitive enough to definitively detect a difference of 5° or 10°. Work by others is consistent with the results observed here, as co-crystal structures of TBP and several TATA variants show that the DNA is bent to the same degree (18). Förster resonance energy transfer (FRET) would be a more sensitive technique than EMSA for determination of the induced bend angles in solution. However, FRET does not incorporate the extensive sequence context of the TATA box, which has been shown to have a significant effect on the stability of the TBP-DNA complex (21, 22), as

the oligonucleotide probes must be kept short so that the distance between the fluorophores is within the range where energy transfer can occur. Although FRET with variant TATA boxes yielded TBP-induced bend angles ranging from 30° to 76° (27), other work using FRET to determine TBP-induced bend angles suggested that TBP bound to suboptimal TATA sequences can yield a population of different conformers (28). It is conceivable that the observed invariance in TBP-induced bend angle (Table 3.2) may be due to selective capture of complexes containing more severely bent DNA within the gel matrix, however, selective enrichment for such conformers was evidently not seen by Starr *et al.* using the same electrophoretic technique (17). We therefore suggest that sequence context of the TATA box exerts a significant effect on the observed bend angle; previous analyses of TBP-DNA complex formation indicate that it is sequence downstream of the TATA box that is particularly capable of affecting properties of the TBP-DNA complex (21, 22).

TFIIIB-induced bending of the TATA box DNA was observed (Figure 3.4), but the slight differences observed with TBP-induced bend angles was effectively "ironed out" by formation of the TFIIIB complex, probably because of the additional bend induced by TFIIIB downstream of the TATA box and upstream of the start site of transcription. In addition, complex formation at the end of the DNA probe is less efficient because the complex may be less stable to electrophoresis (29). The values determined here for TFIIIB-induced bending agree with previous determinations (25). Unexpectedly, the additional bend did not appear to increase the magnitude of the bend in the TFIIIB-DNA complex, as determined by visual inspection of the migrating complexes in the gel and by fitting the data to equations 1 and 2. Since Brf1 and Bdp1 are known to change TBP-DNA contacts (15, 24), and since formation of TFIIIB causes conformational changes in Brf1 and Bdp1 that reveal cryptic DNA-binding domains (13, 30), the

observation that the TFIIIB-induced DNA bend is comparable to that induced by TBP may reflect a conformational change in the TBP-DNA complex to one more amenable to Brf1 and Bdp1 contacts.

The TBP-induced bend angle is not correlated with relative levels of pol III transcription

In contrast to results obtained from similar work on a pol II system (17), the TBP-induced bend angle appears to have little effect on transcription by pol III. There is an indication that pNTS-26T is transcribed less efficiently than the other templates, in agreement with previous results from a pol II system (17); however, this sequence is bent to essentially the same degree as the others (Table 3.2). Despite the lack of difference in the bend angles, there are differences in transcriptional activity.

Given the changes in TBP-DNA contacts induced by formation of the TFIIIB-DNA complex (13, 15, 24, 30), the "ironing out" of TBP-induced bending differences among the probes by formation of TFIIIB (Figure3.4), and our results (31) showing that formation of the TFIIIB complex assembled on DNA with TBPm3 changes the preference for the downstream half of the TATA box from one preferred by TBPm3 or even wtTBP to a sequence that matches the TATA box of the pol III-transcribed *SNR6* gene (see Tables 2.2 and 2.3), it seems that formation of the TFIIIB-DNA complex imposes restrictions on the conformation of the TATA sequence that are more important than the preference of TBP. Molecular dynamics simulations of TBP-TATA complexes have also suggested that for pol II transcription, it is the flexure of the TATA box, not the bend angle, that governs transcriptional efficiency (32, 33).

A further explanation for the differences observed here from the results of the pol II study is the use of a modified pol III gene in the pol III sequence context. Regions upstream of the start site of transcription but downstream of the TATA box have been shown to affect transcription in

SUP4 tRNA genes (19), consistent with the notion that the sequence context has a greater effect on the transcriptional efficiency than the TATA box sequence. Another possibility is the use of full-length TBP in these studies, while the previous work (17) employed TBP (Δ3-60) because it formed greater levels of complex. Although in yeast the non-conserved N-terminus of TBP is not thought to contact the DNA (34), the N-terminus has been implicated in regulation of DNA-binding activity (35), as a region for protein-protein interactions with other proteins involved in transcription (36), and as an inhibitor of transcription *in vivo* (36). Interestingly, the N-terminus of human TBP has been shown to influence the protein-induced bend angle (37).

The sequence-independence of the induced bend at the TATA box observed here may help explain how TBP is interacting with non-canonical TATA sequences *in vivo*. TBP is required at these sequences for transcription to occur (1) and is placed there by other transcription factors, even when no TATA box is present (38). TBP may function on *SUP4* promoters *in vivo* mainly in its capacity to cause conformational changes in Brf1 and Bdp1 (13), as these genes lack a TATA box. While these tRNA genes lack a canonical TATA box, there are A+T-rich sequences present in the region 30 bp upstream of the start site of transcription that are protected by TFIIIB in DNase I footprinting studies (39) and could be considered TATA-like. However, these sequences are not strictly conserved (39) beyond a general A+T-rich nature that is more flexible than random DNA.

Sequence well beyond the terminator affects termination efficiency

Termination from the pNTS templates is unexpectedly inefficient, with the primary transcript corresponding to a read-through product and little transcript terminating at the expected ~100 nt (Figure 3.5). Since the native *SUP4* terminator, which consists of 7 T bases, is present in these templates, it was surprising that efficient termination was not observed, as a T7

terminator is considered optimal for termination of yeast pol III. While determinants of transcriptional termination remain incompletely characterized, previous work has shown that insertion of the strong T5 terminator and 20 bp of downstream sequence that mediates efficient termination of transcription from a yeast tRNA gene into the yeast *SCR1* gene dramatically decreases the efficiency of the terminator (40). Since the native *SUP4* sequence in the pNTS templates ends 5 bases downstream of the terminator, our data suggest that the sequence of the pET5a-Bend parent plasmid may be affecting termination efficiency. Our observation that sequence well beyond the T cluster may affect termination efficiency may also explain the fact that many pol III-transcribed genes contain internal T-stretches that do not lead to transcription termination. This interesting result opens up a new avenue for extensive investigation into the mechanism by which sequence downstream of the terminator affects termination on genes using a T7 terminator.

References

1. Cormack, B. P. & Struhl, K. (1992). *Cell* **69,** 685-96.
2. Schmidt, M. C., Kao, C. C., Pei, R. & Berk, A. J. (1989). *Proc. Natl. Acad. Sci. USA* **86,** 7785-9.
3. Schultz, M. C., Reeder, R. H. & Hahn, S. (1992). *Cell* **69,** 697-702.
4. Kuddus, R. & Schmidt, M. C. (1993). *Nucleic Acids Res.* **21,** 1789-96.
5. Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. (1993). *Proc. Natl. Acad. Sci. USA* **90,** 8174-8.
6. Bucher, P. (1990). *J. Mol. Biol.* **212,** 563-78.
7. Yamamoto, T., Horikoshi, M., Wang, J., Hasegawa, S., Weil, P. A. & Roeder, R. G. (1992). *Proc. Natl. Acad. Sci. USA* **89**, 2844-8.
8. Reddy, P. & Hahn, S. (1991). *Cell* **65**, 349-57.
9. Starr, D. B. & Hawley, D. K. (1991). *Cell* **67,** 1231-40.

10. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature* **365,** 520-7.

11. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature* **365,** 512-20.

12. Kassavetis, G. A., Braun, B. R., Nguyen, L. H. & Geiduschek, E. P. (1990). *Cell* **60,** 235-45.

13. Kassavetis, G. A., Joazeiro, C. A., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S. & Blanco, J. A. (1992). *Cell* **71,** 1055-64.

14. Kassavetis, G. A., Nguyen, S. T., Kobayashi, R., Kumar, A., Geiduschek, E. P. & Pisano, M. (1995). *Proc. Natl. Acad. Sci. USA* **92,** 9786-90.

15. Grove, A., Kassavetis, G. A., Johnson, T. E. & Geiduschek, E. P. (1999). *J. Mol. Biol.* **285,** 1429-40.

16. Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T. & Roeder, R. G. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 1060-4.

17. Starr, D. B., Hoopes, B. C. & Hawley, D. K. (1995). *J. Mol. Biol.* **250,** 434-46.

18. Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T. & Burley, S. K. (1999). *Genes Dev*. **13**, 3217-30.

19. Shaw, K. & Olson, M.V. (1984). *Mol. Cell. Biol.* **4**, 657-65.

20. Librizzi, M. D., Brenowitz, M. & Willis, I. M. (1998). *J. Biol. Chem.* **273**, 4563-8.

21. Wolner, B. S. & Gralla, J. D. (2001). *J. Biol. Chem*. **276**, 6260-6.

22. Faiger, H., Ivanchenko, M., Cohen, I. & Haran, T. E. (2006). *Nucleic Acids Res*. **34**, 104-19.

23. Whitehall, S. K., Kassavetis, G. A. & Geiduschek, E. P. (1995). *Genes Dev*. **9**, 2974-85.

24. Persinger, J., Sengupta, S. M. & Bartholomew, B. (1999). *Mol. Cell. Biol.* **19,** 5218-34.

25. Léveillard, T., Kassavetis, G. A. & Geiduschek, E. P. (1991). *J. Biol. Chem.* **266**, 5162-8.

26. Thompson, J. F. & Landy, A. (1988). *Nucleic Acids Res.* **16,** 9687-705.

27. Wu, J., Parkhurst, K. M., Powell, R. M., Brenowitz, M. & Parkhurst, L. J. (2001). *J. Biol. Chem*. **276**, 14614-22.

28. Powell, R. M., Parkhurst, K. M. & Parkhurst, L. J. (2002). *J. Biol. Chem.* **277**, 7776-84.

29. Braun, B. R., Kassavetis, G. A. & Geiduschek, E. P. (1992). *J. Biol. Chem.* **267**, 22562-9.

30. Kassavetis, G. A., Bartholomew, B., Blanco, J. A., Johnson, T. E. & Geiduschek, E. P. (1991). *Proc. Natl. Acad. Sci. USA* **88**, 7308-12.

31. Tsihlis, N. D. & Grove, A. (2006). *Nucleic Acids Res.* In press.

32. Qian, X., Strahs, D. & Schlick, T. (2001). *J. Mol. Biol.* **308**, 681-703.

33. Strahs, D., Barash, D., Qian, X. & Schlick, T. (2003). *Biopolymers* **69**, 216-43.

34. Khrapunov, S., Pastor, N. & Brenowitz, M. (2002). *Biochemistry* **41**, 9559-71.

35. Perez-Howard, G. M., Weil, P. A. & Beechem, J. M. (1995). *Biochemistry* **34**, 8005-17.

36. Lee, M. & Struhl, K. (2001). *Genetics* **158**, 87-93.

37. Zhao, X. & Herr, W. (2002). *Cell* **108,** 615-27.

38. Gerlach, V. L., Whitehall, S. K., Geiduschek, E. P. & Brow, D. A. (1995). *Mol. Cell. Biol.* **15**, 1455-66.

39. Kassavetis, G. A., Riggs, D. L., Negri, R., Nguyen, L. H. & Geiduschek, E. P. (1989). *Mol. Cell. Biol.* **9**, 2551-66.

40. Braglia, P., Percudani, R. & Dieci, G. (2005). *J. Biol. Chem.* **280,** 19551-62.

CHAPTER 4

SUMMARY AND CONCLUSION

The TATA-binding protein (TBP) is universally required for transcription by all three eukaryotic RNA polymerase (pol) systems (1), and the sequence of the core DNA-binding region is conserved from archaea to humans (2). TBP interacts with the DNA in a unique manner, binding in the narrower minor groove (3) and causing extreme deformations to the double helix (2, 4), while still maintaining Watson-Crick base pairing (Figure 1.7). The large bend angle induced in the TATA box by TBP brings proteins involved in transcription closer together to facilitate their interaction (5 and Figure 1.8). Several observations have led to the evaluation of this phenomenon in the pol III system: i) the reported sequence dependence of the magnitude of this bend angle in a pol II system and its correlation with transcriptional activity (6); ii) the similarity of pols II and III (see Table 1.1); iii) the involvement of TBP in both systems; and iv) the additional bend induced between the TATA box and the start site of transcription by formation of the RNA pol III recruitment factor complex (TFIIIB) (7).

Formation of TFIIIBm3 Imposes a Sequence Preference on TBPm3

An iterative *in vitro* selection was performed using TBPm3 to assemble the TFIIIB complex on a *SUP4* tRNA promoter, modified to contain a mutant TGTA box, and determine the sequence preference of this complex for the downstream half of the TGTA box. TBPm3 was purified and its binding parameters characterized (Table 2.1), which verified it dissociated from DNA with first-order kinetics and bound DNA with second-order kinetics, similar to wtTBP (8). The *in vitro* selection was performed on TBPm3 alone to determine its sequence preference at the downstream half of the TGTA box. Despite reaction conditions meant to make the selection more stringent, only a modest sequence preference for non-C bases was observed (Table 2.2).

Appearance of a non-A/T base at position N1 was unexpected, but may reveal a conformational change in TBPm3 that allows for this lack of stringency. That TBPm3 bound properly to a sequence representing the favored bases at each position, was verified by electrophoretic mobility shift assay (EMSA, Figure 2.8) and MPE-Fe(II) footprinting (Figure 2.9), in spite of this sequence not being present in the selected clones. The footprinting showed enhanced cleavage at positions -18, -19, and -31, consistent with DNA made more accessible due to flexure at these sites, and identical to these enhanced cleavage sites observed in the wtTBP footprint (7-9).

More interesting was the changes caused to the sequence preference by formation of the TFIIIB complex assembled with TBPm3. In this case, a strong preference emerged for a sequence that was different from that preferred by wtTBP or TBPm3 alone. In fact, the selected sequence, present in almost all of the clones with a TGTAA box, was identical to that of the U6 gene (Table 2.3), which contains a TATA box. This indicates that the selected sequence favors entry of Brf1 and Bdp1 into the TFIIIB complex. Given the malleability of the T•A step (10), we suggest that flexure at the downstream half of the TATA box promotes assembly of TFIIIB.

Sequence Context Effects on Protein-Induced Bend Angles

Previous work on a pol II system (6) showed that increased TBP-induced bend angles correlated positively with increased levels of relative transcription on a set of TATA boxes of different sequence. TBP-induced bend angles were determined using circular permutation to move the TATA box with respect to the ends of the DNA, EMSA was used to separate the bound DNA from the free DNA, and anomalous migration due to the induced bend served as the basis for estimation of bend angles. As seen in Tables 3.2 and 3.3, sequence at the TATA box has almost no effect on the protein-induced bend angle at these sequences. FRET experiments by

others (11) have shown that TBP bound to a variant promoter (TATAAACG) exhibits decreased bending and relative transcription activity. This issue remains unresolved, however, as X-ray structures show that the bend angle is unaffected by the TATA box sequence (12), while FRET studies show that TATA box sequence causes different degrees of bending (13). Poor TATA boxes were also shown by FRET to form an ensemble of TBP-DNA conformers (11), some of which are perhaps more amenable to capture by crystallization or EMSA.

Sequence Well Beyond the Terminator Affects Termination Efficiency

In vitro transcription reactions assembled with purified components (TBP, Brf1, Bdp1, pol III) on pNTS and its derivatives show several RNA products, the largest and most abundant of which corresponds to a read-through transcript (Figure 3.5). The band observed is RNA, as confirmed by its removal upon addition of RNase A. The band is also dependent on the TATA box, as an *in vitro* transcription reaction using pET5a-Bend, which does not contain the *SUP4* gene, as the template eliminated the production of RNA transcript. Quantitation of the most prominent transcript showed that there are differences in relative transcription despite no difference in the bend angle (Table 3.3). The levels of relative transcription observed here are also distinct from those observed in the study using a pol II system (6); therefore, it may be that the flexure or dynamics of the TATA box sequence governs transcription efficiency.

The observed inability of transcription to terminate efficiently (Figure 3.5) is notable, as the pNTS templates contain the native *SUP4* T7 terminator and 5 base pairs of downstream sequence. This suggests that sequence well beyond the terminator may be affecting proper termination of the transcript. When a strong terminator—and sequence 20 base pairs downstream—was removed from a yeast tRNA gene and placed in the yeast *SCR1* gene, the

ability of the terminator to terminate transcription properly was compromised (14), suggesting that termination by pol III requires more than the well-characterized string of T's.

Future Work

To assess sequence context effects on TBP-mediated bending, FRET studies with short oligonucleotides may be performed to compare to the TBP-induced bend angles determined by EMSA. It bears mentioning, however, that since the oligonucleotides used for FRET are short, the effects of the sequence context on bending will be limited. The unexpected weakening of the T7 terminator used in pNTS that derives from substituting downstream sequence indicates a previously unappreciated dependence on sequence other that the stretch of T's thought to suffice for termination. Mutagenesis of the sequence beyond the terminator in pNTS can serve as a means to explore the effects of sequence context beyond the terminator on transcription.

References

1. Cormack, B. P. & Struhl, K. (1992). *Cell* **69,** 685-96.
2. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature* **365,** 512-20.
3. Starr, D. B. & Hawley, D. K. (1991). *Cell* **67,** 1231-40.
4. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature* **365,** 520-7.
5. Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T. & Roeder, R. G. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 1060-4.
6. Starr, D. B., Hoopes, B. C. & Hawley, D. K. (1995). *J. Mol. Biol.* **250,** 434-46.
7. Grove, A., Kassavetis, G. A., Johnson, T. E. & Geiduschek, E. P. (1999). *J. Mol. Biol.* **285,** 1429-40.
8. Grove, A., Galeone, A., Yu, E., Mayol, L. & Geiduschek, E. P. (1998). *J. Mol. Biol.* **282**, 731-9.
9. Hoopes, B. C., LeBlanc, J. F. & Hawley, D. K. (1998). *J. Mol. Biol.* **277**, 1015-31.

10. Goodsell, D. S., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1994). *J. Mol. Biol*. **239**, 79-96.

11. Powell, R. M., Parkhurst, K. M. & Parkhurst, L. (2002). *J. Biol. Chem.* **277,** 7776-84.

12. Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T. & Burley, S. K. (1999). *Genes Dev*. **13**, 3217-30.

13. Wu, J., Parkhurst, K. M., Powell, R. M., Brenowitz, M. & Parkhurst, L. J. (2001). *J. Biol. Chem*. **276**, 14614-22.

14. Braglia, P., Percudani, R. & Dieci, G. (2005). *J. Biol. Chem.* **280,** 19551-62.

VITA

Nick Tsihlis was born in Houston, Texas, where he attended Scarborough High School. In the fall of 1995, he entered Rice University, and received his Bachelor of Arts in biochemistry and cell biology in January 2000. He enrolled in the biochemistry master's program in the Department of Biological Sciences at Louisiana State University in the spring of 2000 and joined the laboratory of Dr. Anne Grove. In the spring of 2001, he entered the doctoral program, where his primary research involved investigation of TBP-DNA contacts in the *Saccharomyces cerevisiae* RNA polymerase III recruitment factor TFIIIB and their effects on transcription. As a graduate student, he taught laboratory courses in freshman general biology, senior biochemistry, and sophomore microbiology. In addition, he served as an instructor in a summer program, funded by the Howard Hughes Medical Institute, aimed at teaching area high school teachers how to perform simple and illustrative molecular biology experiments with their students. Nick will complete the requirements for the Doctor of Philosophy degree in biochemistry in December 2006. In July 2006, he joined the laboratory of Dr. Melina Kibbe in the Department of Vascular Surgery at the Northwestern University Feinberg School of Medicine in Chicago as a post-doctoral fellow.